025 R35n
Reichman $8.00
Notched cards

62-12048

373+21=394

THE STATE OF THE LIBRARY ART

Publication of
THE GRADUATE SCHOOL OF LIBRARY SERVICE
Rutgers - The State University

THE STATE OF THE LIBRARY ART, edited by
Ralph R. Shaw

Volume I, Parts 1 and 2 (in one volume) $8.00
Cataloging & Classification, by Maurice F. Tauber;
Subject Heading, by Carlyle J. Frarey

Volume I, Parts 3 and 4 (in one volume) March, 1961
$8.00 Classification, by Maurice F. Tauber and
Edith Wise; Gifts & Exchanges, by Donald E.
Thompson

Volume II, Parts 1 and 2 (in one volume) $5.00
Training Laymen in Use of the Library,
by George S. Bonn; Bibliographical Services,
by Margaret S. Bryant

Volume II, Part 3 January, 1961 $8.00
Charging Systems, by Leila H. Kirkwood

Volume III, Parts 1-3 (in one volume) $5.50
Library Buildings, by Ralph E. Ellsworth; Shelving,
by Louis Kaplan; Storage Warehouses, by Jerrold Orne

Volume IV, Parts 1-5 (in one volume) March, 1961
$8.00 Notched Cards, by Felix Reichman; Feature
(Peek-a-boo) Cards, by Lawrence S. Thompson;
Punched Cards, by Ralph U. Blasingame, Jr.;
Electronic Searching, by Gerald Jahoda; Coding in
Yes-No Form, by Doralyn Hickey

Volume V, Part 1 $5.00
Production and Use of Micro-Forms, by
Reginald R. Hawkins

Volume V, Part 2 $5.00
Reading Devices for Micro-Images, by Jean Stewart
and others

Volume V, Part 3 $8.00
Full-Size Photocopying, by William R. Hawken

Metcalf, Keyes D. ed.
Studies in Library Administrative Problems. $5.00

Distributed by
The Rutgers University Press
New Brunswick, New Jersey

the state of the library art

edited by Ralph R. Shaw

volume 4, part 1

Notched Cards

by Felix Reichman

volume 4, part 2

Feature Cards (Peek-a-Book Cards)

by Lawrence S. Thompson

volume 4, part 3

Punched Cards

by Ralph Blasingame, Jr.

volume 4, part 4

Electronic Searching

by Gerald Jahoda

volume 4, part 5

Coding in Yes-No Form

by Doralyn J. Hickey

Graduate School of Library Service
Rutgers - The State University
New Brunswick, N.J. 1961

L. C. Card No. 60-16772

This volume is one of a series of studies in librarianship published by the Graduate School of Library Service * Rutgers - The State University and distributed by the Rutgers University Press

Preface

Pierce Butler planted the seed of this work some thirty years ago when he attempted to persuade me to follow my translation of Georg Schneider's "Theory and History of Bibliography" with an English rendering of Milkau's "Handbuch der Bibliothekswissenschaft." As we discussed this in his cubbyhole at the University of Chicago it became increasingly apparent that for all the excellence and usefulness of the Milkau, a version in English would require such wide revision and additions as to constitute a completely new work, and the project was, I thought, abandoned. Over the years, however, it remained a subject of, perhaps sometimes sub-conscious, dreaming, thinking, and occasionally some work.

When the Council on Library Resources was organized its need for information on the state of the library art resulted in almost spontaneous germination of Pierce Butler's long dormant seed. This resulted in a grant to the Graduate School of Library Service at Rutgers for preparation, under the direction of the undersigned, of a review of the status of our current knowledge of librarianship. From the inception of this project it was recognized that the initial grant could cover only somewhat less than half of the vast field, and there are many areas of librarianship still to be covered.

An advisory committee helped in the design of the program and in determining priorities for treatment of the various aspects of the field.

The advisory committee included: Dr. Julian H. Bigelow, Institute for Advanced Study at Princeton, New Jersey, Mr. Verner W. Clapp, President of the Council on Library Resources, Mr. Donald Coney, Librarian of the University of California, Mr. J. W. Kuipers, Appa-

ratus research and development in Itek Corporation, Dr. Robert D. Leigh, Dean of the School of Library Service at Columbia University, Dr. Lowell A. Martin, Dean of the Graduate School of Library Service at Rutgers, Professor A. J. Riker of the College of Agriculture at the University of Wisconsin, Dr. Melville J. Ruggles, Vice President of the Council on Library Resources, Mr. A. N. Sears, Vice President of Remington Rand, and Dr. Eugene H. Wilson, Vice President of the University of Colorado.

Since the volumes now completed and being prepared for publication cover perhaps forty percent of the entire range of librarianship and bibliography, the plan of publication is in the familiar "Handbuch" form so as to permit the addition of volumes as financial resources and dedicated writers become available.

In preparing these for publication it appeared best to permit some variations in style from one volume to another rather than to devote a large percentage of our resources to achieving a standard style for all.

The general pattern followed in these studies consists of a survey of the published and unpublished literature of each facet of the field. In this survey, as a first step, each compiler attempts to summarize what the literature says with a minimum of redundancy but without editorial comment. Each statement is accompanied by a footnote so that investigation in depth can be conducted when necessary, but for most purposes, if we have adequately performed our primary task, it should be unnecessary to search the literature for information on the topics covered. A second step, provided in most cases, is examination of the evidence provided to support each allegation or statement in the literature and an indication of whether that particular bit of "the art" is empirical or the extent and reliability of the objective data provided to support it.

This pattern of presentation is modified in a number of the parts of the series since, except for the sake of external uniformity, it would serve little purpose to repeat substantially every statement in this second part followed by the words, "no objective evidence."

It was initially planned to present each summary in the historical perspective of the development of its field but this method of treatment was found unsuitable for some of the subjects. In some cases, such as the study of "Reading Devices for Micro-Images," both historical and topical treatments are presented; in others the treatment is historical only, and in still others primary emphasis is given to topical treatment. Here as in the review of evidence, suitability for each topic is given priority over external uniformity of the set.

While a few of these reports were prepared by staff assistants, as indicated by the list of titles and research staff, we were fortunate in enlisting participants of outstanding authority and reputation in most of the subject areas treated. In most cases the resultant first draft was read by one or more additional specialists in the field and in this work we had the assistance of such well known authorities as: Hubbard W. Ballou, Ralph H. Carruthers, Ralph Esterquest, Robert A. Fairthorne, John Fall, Charles F. Gosnell, Lutz Helbig, Laurence Kipp, Alfred H. Lane, Chester Lewis, Calvin N. Mooers, Robert H. Muller, Maurice F. Tauber, Lawrence S. Thompson, and others.

This is the concerted work of many hands, and many more than those listed above have helped in almost countless ways. Such value as may reside in this series is the result of professional contributions of high order by many people. Final responsibility for the plan of work, selection and supervision of research staff, revision of manuscripts as well as editing them and production of the volumes rested solely on the undersigned and he accepts full responsibility for the imperfections to be found in this series. It is hoped that this review of the library art will, on balance, be found to make a useful contribution and that it may some time be carried forward to cover the whole field of librarianship.

Ralph R. Shaw

Table of Contents

Volume Four   Part One

NOTCHED CARDS

by

Felix Reichmann

The function of a library is not confined to the acquisitions and preservation of material. Though many librarians have bibliophilic interests and are intent on enriching their collections, yet the ultimate goal of the library is use. The modern conception of technical services in libraries regards acquisition as an integral part of the preparation process and considers its task fulfilled only when the book has passed the three stages: selection, acquisition, cataloging, and has been placed at the disposal of the reader.

Much thought has gone into the problem of how to organize library material for use, and different possibilities have been carefully examined. The conventional dictionary catalog, its division into an author and a subject file, the classified arrangement of catalog cards, the formation of smaller bibliographical files destined for a selected clientele (technical reports in special libraries, imprint files in Rare Book Departments, etc.) and finally the physical arrangement on the shelves according to a subject classification or by current number--all have been described and their specific advantages recognized.

All these bibliographical techniques have certain properties in common: 1. The method of organization (the input of information) determines the retrieval possibilities. For example, an author catalog can only tell which titles of a given author are available, and an imprint file will provide only information as to place and publisher or as to date of publication. 2. The nature of the information to be given must be envisaged from the outset; no provision can be made for new categories. 3. An excessive number of cards would be necessary to bring out all possible combinations of subject aspects, e.g. a title with five descriptors would need 120 cards ($5! = 120$). Cross-references are only a partial remedy for this shortcoming. 4. The cards

must be kept in a specific sequence. 5. It is extremely difficult to substantiate a negative answer, which in scholarly research is often as important as a positive one.

The complexity of modern research and the awareness of the multitude of subject relations which may be significant in future scholarly endeavours impel librarianship to search for new methods of bibliographical control which would augment the conventional techniques and permit a greater flexibility in the retrieval of information (1). A simple practice to increase the efficiency of a cardfile without adding a new subject heading or being forced to inspect every single card, is the use of little colored flags or notches punched out from the upper edge of the card to signal specific information. The latter device (triangular notches made with the Copeland-Chatterson single hole punch) is employed at the classified catalog of the Institute of Cancer Research (London) to identify titles which have "a considerable and easily usable bibliography" (2). In this instance punches are auxiliary to conventional cards, the punched cards reverse the roles and assign the primary function of information retrieval to holes and notches (3).

Punched cards as memory-storage of patterns were employed to control the operations of the loom invented by Joseph Marie Jacquard in 1801. Herman Hollerith adapted the basic idea for the construction of a tabulating machine for the U. S. Census Bureau in 1886. The first patent for a marginal punched card was granted to H. P. Stamford (1896, No. 564, 117), subsequent patents were granted to W. M. Stretch (1907, No. 867, 618), and E. C. Molina (1914, No. 1, 083, 456). Independent of these still rather primitive inventions a notched card was used to record information gathered at a hookworm survey in Brazil in 1920. The decisive step to produce a generally applicable punched card was taken by Alfred Perkins. Mr. Perkins resorted to marginal punches to sort tickets used by the Dunlop Rubber Company in Birmingham. The printer of these tickets, Copeland-Chatterson, obtained the inventor's permission to apply for an English patent on a royalty basis and became the pioneer in the commercial distribution of marginal

punched cards.

Mr. Perkins received an American patent in 1925 (U.S. Patent 1,544,172) and sold the U.S. rights to the McBee Corporation in 1932. In subsequent years scores of patents both in this country and in Europe have been granted to improve upon or to vary the Perkins invention (3a).

The characteristic features of the marginal punched card are: 1. The cards have one row (or multiple rows) of pre-perforated holes around the edges. 2. The center of the card remains free for conventional media of information (typing, design, microfilm, etc.). 3. All retrieval aspects of one title are recorded on one card, thus eliminating the necessity of multiplying cards. 4. Combinations of aspects, although not perceived when the file was set up, can be searched for. One may generalize that punched cards reduce the conventional catalog in space but extend its usefulness in time. The cards can show both logical sums (a+b, also called alternation as it also signifies a or b), and logical products (ab). 5. Cards can be filed at random.

The versatility of the card demands great care and exactness in the formulation of the information to be transmitted. The literature of the field reiterates the stern advice: "No device, simple or complicated, can compensate for thoughtlessness in the analysis of information or for sloppiness in the use of terminology" (4).

In art history it has been observed that the introduction of new material necessitated the adaptation of new techniques and thus created a new style (5), similarly new bibliographical techniques often work more efficiently if they are based on new semantic formulations which have been constructed in keeping with the characteristics of the new tool.

This statement, however, is not universally accepted. Many American designers of punched-card equipment suggest to their customers the use of "tailor made" descriptors often sharply deviating from library terminology, but Bloomfield cautions "Let's not reject conven-

tional subject-headings" (6). Haykin advocates the construction of "classifications chiefly designed for information retrieval" (7) and Stroem believes that multidimensional schemes may be preferable to conventional ones (8). Conversely, many European librarians write most favorably of their experiences with the Universal Decimal Classification. Ruston considers it "Easily applicable, ... however, only figures can be used; all other symbols have to be eliminated" (9). Grobe reports "Excellent results" (10) and Weigelin praises the punched-card file of the eye clinic in Bonn which has applied the UDC to integrate the material from different eye hospitals (11). Finally the International Federation for Documentation fairly endorses the use of this system in the conclusion of the chapter on Decimal Classification (12). "The UDC is a suitable classification basis for mechanical selection in documentation, without any restriction for edge punched cards, but with some restrictions for surface punched cards. It is distinguished for its encyclopedic structure, its numerical basis, its international character and above all, for the test of practical use which it has survived. All bases for coding of content of articles are available. Future experiences with mechanical aids can easily be worked in with the UDC."

Punched cards transmit information through holes and notches and not through letters and numerals. The translation of the conventional medium into the language of the punched card is called coding. One of two things can be done to the prepunched hole of a marginal punched card; it can be left intact, or it can be enlarged to reach either the edge of the card or the next hole. We have thus a binary code where every digit (letter, word) is either 0 or one. It is by no means a new invention to convey information through a two-state code. Cryptography has used it often, Francis Bacon employed it in his biliteral code and the dots and dashes of the Morse Code are well known (13).

It is important to remember that the hole, if left untouched, does not represent any value, although the explaining label is printed alongside. It becomes meaningful when it is notched and the hole thus extends to

the outer edge. If the sorting needle is introduced into the hole, the notched cards will fall from the needle, these being the cards which have been selected. Two additional operations are feasible with cards which have a double row of holes. The hole in the inner row can be connected with a notched hole and thus reach the edge (deep punch), or two holes can be connected without reaching the edge, (intermediate punch). In the latter case the card will not fall from the needle but only drop 1/4 of an inch. A second needle has to be introduced into a guide hole and the first withdrawn to have the card drop out completely. A third variation is theoretically possible. If two holes in the inner row can be connected horizontally, the card would not drop but slide sideways (14). Cards with three rows add other possibilities of punching. Cards with multiple rows (more than three) are rarely used.

The single hole which corresponds to a digit (letter, word, etc.) is called a code position. A number of code positions, generally four to six, all pertaining to a single subject or classification, can be combined into a code field. Two or more code fields related to the same subject can be joined in a code section.

The simplest method of coding which assigns only one meaning to every hole is called direct coding. The question which has to be answered through the medium of one sorting needle is a simple one: yes or no. The hole left intact is "no", the notch is "yes" (15).

The capacity of the edge punched cards for information storage is very great indeed. A card with a single row of twenty-six holes permits a variety of combinations equal to two to the power of twenty-six. Every hole permits one of two operations and as the card has twenty-six holes approximately seventy million combinations are possible, in the case of fifty-two holes the possibilities of variations would be four thousand billion. That means that in a library of seventy million titles, each title could be identified with direct coding by using a punched card file with twenty-six holes to the card. The characterization, however, would consist in a variety of twenty-six conceptions. Obviously this is a

purely academic proposition. More important from the practical point of view is the serious disadvantage that the sum total of aspects to be coded cannot be higher than the number of holes on a single card. This method, however, is the simplest coding technique and the easiest way to retrieve information. Furthermore, if the questions are formulated in keeping with the coding structure, there can be no unwanted cards. It is not always the fastest system; if the digits 0-9 are coded directly nine operations with the needle will be necessary to sort a pack of cards numerically from zero to nine and every card would be handled on the average of 5 1/2 times (16).

Slightly more sophisticated is the numerical sequence code generally represented by the code field: seven, four, two, one. This field covers all the digits from zero to nine. For zero the field is left intact, for three both two and one are notched, for five both four and one are notched, and so on. Only four needling operations are necessary and the number of cards to be needled would be about four times the size of the pack. As the notches are used in combinations, (four and one have to be combined to give five) this method is also called combination code. It is also possible to code the digits 0-14 in a four hole field. A code section can be arranged where the first field covers the units, the second the tens, the third the hundreds and so forth. It is interesting to note that the numerical sequence runs from right to left.

By adding a fifth hole to the field this method can be adapted to serve as a simple alphabetical code. In this case the holes represent the letters of the alphabet instead of digits, A = one, B = two, C = two plus one and so on for the first thirteen letters, a-m. To identify the second thirteen letters, n-z, the fifth hole marked n-z, is notched with N = one, O = two, P = two plus one, etc. Only one number (or one letter) can be coded in any one field. One needle is being used and as the single hole has a combination of meanings, an unequivocal selection of a card is not possible, however it is a most convenient method to order the cards in a se-

quence.

A better alphabetical code has been made by Cox, Bailey and Casey (17). A code field of five positions is marked with the letters O I E C B. All letters of the alphabet can be coded, most of them in combinations of two to four holes. Five letters need one position, ten letters require two, nine letters three, and three letters four positions. "A" is identified by leaving the field intact. There are twenty-eight symbols, because M is subdivided into three characters. This code too is applicable for sequence sorting only.

If all cards of a certain group must be selected without having too many unwanted cards (the so-called noise-effect) a selector code has to be employed using two needles in a field of six holes. The most common arrangement is SF seven, four, two, one, zero. SF (single figure) is notched together with seven if this digit is to be selected, for the selection of eight, holes seven and one are notched; however for coding zero, this hole alone is notched.

This method is also necessary if the number of conceptions to be coded exceeds the number of holes. A six-hole field, using two positions to code one aspect, would give the choice for one out of fifteen possible conceptions: if three holes are combined to code one aspect, the choice increases to twenty. The mathematical rule is that the maximum number of aspects is given if the number of positions used for one code is half the number of holes in one field. However, the number of possible choices has to be weighed against the ease of retrieval. A code with three choices needs three needles where a two-position code needs only two. Only one aspect can be coded in any one field.

In the triangle code (field of five holes) the ten symbols (0-9) are not printed parallel to the holes but arranged in a triangle or pyramid. The two holes are notched whose diagonal columns intersect at the digit or letter which has been selected. All the symbols can be used directly from the card without remembering the combinations necessary to form the letters or digits;

moreover the field has been reduced from six to five holes.

If the overall number of subjects to be coded is very large and more than one aspect should be recorded in any one field, superimposed coding has to be employed. Each concept is coded by notching two or more holes, completely disregarding the fact that some of the notches may have been used already in another combination to identify another descriptor. A certain "noise effect" (unwanted cards) is therefore unavoidable, however the number of wrongly selected cards will be kept to a minimum if the combinations are selected at random. The reasons for this efficient result are the significant statistical possibilities of random selection. It has been calculated that optimum results will be obtained if forty-six per cent of the positions in any given field are being used. Because of the overlapping code, actually only thirty-seven per cent of the holes available will be notched. Translated into practical operations this means that twelve letters can be coded in an alphabetical field of twenty-six positions.

Calvin N. Mooers has made important contributions to the mathematical theory of coding based on random selections (18). His method called Zatocoding, uses a card with forty holes, not broken up in sub-fields. Sixty-nine per cent of the positions can be utilized; because of overlapping, however, only fifty per cent are actually punched. The single code is based on a four-position pattern. Therefore, in a card with forty holes, seven descriptors can be recorded ($\frac{40 \times 0.69}{4} = 7$). A larger card with seventy-two positions can identify twelve conceptions.

Gilbert proposed a modification of the Zatocoding principle which he called orthographic single-field superimposed code. Instead of random numbers, pairs of letters taken from the spelling of the aspects constitute the code. A card with fifty-five positions is recommended, which can take care of 676 possible English letter pairs, ($26^2 = 676$); twelve to thirteen descriptors are permissible per card (19).

Regardless of the coding system which has been adapted, the coding capacity of the card increases in proportion to the number of holes available. It would be impracticable to enlarge the size of the card beyond a certain point; therefore card manufacturers resort to a double row (or multiple rows) of pre-punched holes. Fairly common is a double row arrangement for the numerical code seven, four, two, one. A "deep punch" enlarging the hole in the second row up to the edge of the card identifies the selection of a single digit (e.g.: seven), the holes in the outer row will be notched if a combination of holes is necessary (seven and one for eight) the same as in a single-row arrangement. The triangular code too is often employed with a double row of holes (20).

An interesting combination of a double and triple row has been worked out by the Oak Ridge Laboratory of the Atomic Energy Commission (21). The card (E-Z Sort) has 366 positions arranged in triple rows at top and bottom and double rows at the two sides. The top edge has two alphabet fields of twenty-three columns each in triple rows designated for subject indexing. The bottom edge has five fields in triple rows. One alphabet field in twenty-three columns, three fields with six columns for author's name and one with three columns (for bibliographical information); each side has six fields, four columns in double rows each. These fields are double printed, either in the numerical code seven, four, two, one, or as a code section with a twenty-four-letter alphabet. They will identify the classification system or serve "any other use desired". The coding system is direct coding; subjects are coded by the first four letters. E-Z Sort offers a thirty position code field arranged in triple rows and marked with the digits 9-0. All digits from 000-999 can be identified with direct coding and selected with three needles. The outer row is notched for the units, the second row is deep-punched for the tens and the third row designates the hundreds. Three positions are needed to code one digit and three separate control holes identify the repetition of a digit in a given number. The control holes are marked H for repetition in the hundreds (883) and T for duplication in the tens (833).

A multiple-row arrangement provides a very great storage capacity. Hardy has calculated the possibility of transferring the London Telephone Directory to a set of pre-punched cards. Both the subscriber's name and his telephone number consisting of three letters and four digits could be recorded on a triple-perforated card. He concludes that "a very fair telephone directory on punched cards could be assembled which could be used for finding names from numbers or numbers from names" (22).

Crosz contributed an interesting probability calculation for a punched card with multiple-column arrangement. A sub-field with twenty-four positions arranged in three columns with eight rows or eight columns in three rows would have a storage capacity of 2024 aspects (23).

From the point of view of information retrieval one has to keep in mind that the needling time increases with the growth of information storage. Furthermore, the combination of different methods of punching (deep punch, intermediate, shallow) make the retrieval time-consuming, cumbersome and difficult.

It is impossible to give any overall preference to any one of the four major methods of coding as every system has "its own peculiar advantages and disadvantages" (24). If the number of positions available is equal to or greater than the sum of the aspects to be coded, direct coding is the simplest and most efficient technique. If the number of conceptions which are needed increases beyond the number of holes, one has to resort to a combination of notches as provided by the sequence and selector codes. These methods are restricted to the selection of one aspect per field and to mutually exclusive aspects to be assigned to different fields or sections. Superimposed coding provides for a large number of concepts which are not mutually exclusive. Information retrieval, however, becomes more time consuming when the coding methods increase in complexity.

In addition to edge-punched cards two other forms

of cards can be sorted manually and must be briefly mentioned in this review. The slotted cards have many features in common with the cards which have been described above (25). They are also called punched cards with central punching (cartes a pre-perforations centrales, Schlitzlochkarten, or Feldlochkarten.) They are coded by punching out the space between two perforations and thus producing a slot, either horizontal or vertical, similar to the "intermediate punch." Most such cards have between two hundred and three hundred holes and are coded in the direct or combination method. Sorting operations often require multiple needles and as freehand needling would be cumbersome, if possible at all, most systems use especially designed but comparatively simple mechanical contrivances for information retrieval.

The American Findex card reserves the upper part of the card for typed or printed information. About seventy per cent of the card space is pre-punched with round holes which are arranged in ten to thirteen columns, every column having ten to fourteen perforations. Two holes slotted in vertical direction form one code position. The positions are numbered from the bottom left corner. A typical Findex card has hundred coding positions arranged in ten columns with eleven perforations each. Both direct coding and a combination code can be employed; a two positions code (using two needles) needs ten positions for forty-five aspects, a three positions code can place one hundred and twenty aspects on ten positions. Retrieval is performed with the help of a selector.

The German card produced by Allform provides slightly more than the upper half of the card for conventional transcription of information. Coding is performed by vertical slots within thirty-four columns of ten round holes each. The cards are also adaptable both to sequence coding and superimposed coding.

Three French systems have horizontal slots and depart more radically from the features of the punched cards which have been reviewed so far.

Selectri leaves the left portion of the card blank for conventional recording. The coding field consists of eighteen to twenty-two columns each with twelve round perforations.

Détectri is similar but has a coding field of seven to thirty-eight rows of rectangular perforations, generally ten to the column. There is ample space on the left part of the card for uncoded information.

The characteristic features of the Dequeker cards are their formats and the Selector. The lower part of the card which contains the coding field is narrower than the upper part left for conventional recording. The coding position is, as in the two other French systems, a horizontal slot, in this instance made between twenty-five columns, each having five round perforations. The Selector is especially interesting. The cards are not in direct contact with one another but separated by "spacers." After the needles have been inserted, the Selector with the spacers moves sideways to the left. The spacers move the cards which have been slotted and the selected cards can be consulted in the Selector and need not be removed.

The second system, Punched Cards with Visual Punching, is based on an entirely different principle (26). Instead of recording each title with its aspects on one card and selecting the titles desired by needling the file of "title cards," this system does exactly the opposite. It allots to every subject one card with a pre-printed grid and punches small holes in the positions which correspond to the serial numbers of the titles pertaining to the given subject. Information retrieval is made by superimposing a number of subject cards; the holes which have been punched in all the subject cards give the serial numbers of the title which represent the desired combination of aspects. This system can thus be called a subject-card system in contrast to the usual title or document card system.

Holmstrom prefers the designation "superimposed coincidentally punched cards" (27). Wildhack and Stern (National Bureau of Standards) speak of "Optical coinci-

dence subject cards" (28). In America it is best known as Peek-a-Boo, in Europe as the Batten-Cordonnier system. (The official nomenclature in France is fiches superposables; in Germany, Sichtlochkarten).

The first systematic use of the subject card system was made by H. Taylor in 1915, who applied it to the identification of birds (29). Five years later H. Soper received an American patent for the use of subject cards in compiling statistical data, and in the same year (1920) C. J. Gray described its application for the identification of minerals in the Transactions of the Geological Society of South Africa. A French patent for searching personnel files by this method was granted to Bourgeaud and Liber in 1923. The idea was further developed by J. Cordonnier and formed the basis of the most important French aspect cards. The first application to literature searching was made by the German librarian R. Preddek in 1930. Preddek's serial numbers referred to his collection of metal plates (Adrema) from which he printed the titles of the appropriate papers. The best known English example of this method is W. E. Batten's control of patent files (30).

The capacity of the majority of available cards ranges between 500 and 20,000 serial numbers. Batten uses an IBM card with 960 positions, the French Sphinxo card has 1000 and Cordonnier's Selecto offers a variety of cards with up to 20,000 positions. The German manufacturers, Ekaha and Allform have models with a capacity varying from 1860 to 7000 numbers. The Dutch "Delta card" has a lozenge pattern placing 10,000 numbers on the intersection of the lines (31). The Termatrex card distributed by Jonker Business Machines has 10,000 and 40,000 positions respectively.

The Office of Basic Instrumentation in the National Bureau of Standards is developing the most far-reaching application of this method (32). Its cards have no preprinted grid and a transparent overlay has to be used with a millimeter grid which permits the punching of up to 18,000 positions. The card stock is of venylit plastic which so far has met all the requirements of "durability, dimensional stability, opacity, suitability

for typing and punchability." Punching errors can be easily mended: a piece of paper card stock is punched into the hole of the plastic card. "No cementing appears necessary; the fibrous insert spreads sufficiently to look itself firmly in place" (33).

The method has several great advantages. By inserting a new aspect card, a new subject can be added with ease as there is no code which has to be re-arranged. It is only necessary to note on the subject card that it had been introduced from a certain serial number on. The system will produce the desired numbers very fast, because only a small number of subject cards has to be superimposed instead of searching in the entire file of title cards. It is necessary, however, to add a second step and to locate the selected titles in their numerical arrangement.

The system works well in small and middle sized collections, no report on the efficiency of its control over large holdings has been available so far. Gagarin uses it in the Gmelin Institute and praises "the simplicity and elasticity of the system" (34). The Office of Basic Instrumentation uses this method as its principal means for searching, and expresses satisfaction with the results. The files contain 1000 subject cards and control about 25, 000 titles. The standard production per day is about 2000 punches (about 200 documents). This, of course, does not include the much more expensive time of the document analyst.

Like most of the hand-manipulated systems, Peek-a-Boo is at its very best if applied to a small collection. It could very well be an almost ideal method of controlling a private collection of reprints. Rothschuh, for instance, reports most favorably on its usefulness in organizing a collection of over 4000 reprints in the field of physiology with 350 aspect cards (35).

The expansibility of the system is considerable: Kistermann is working on a combination of marginal punched cards and Peek-a-Boo, and Stern's Microcite will combine abstracts and aspect cards (36). This de-

vice, which is being explored by the Office of Basic Instrumentation, consists of a film matrix of greatly reduced (30:1) abstracts. The place of the abstract on the film coincides with the position of the corresponding serial number on the cards. With the help of a light diffuser and a microscope the abstract can be read or can be magnified and projected on a screen. There is the further possibility of printing the abstract on a sensitized card. At the moment a matrix is used which contains 1000 microabstracts. The Office is also considering the possibility of a subject card 42" x 22" with a capacity for 500, 000 positions.

The same principle (aspect card and no title card) forms the basis of the Uniterm system or Coordinate Indexing developed by Mortimer Taube and Associates (37). The serial number is not punched in on a predetermined position but is written on the card according to a simple ingenious device. The Uniterm card is divided into ten columns marked by the digits 0-9. The numbers are entered in these columns according to their final digit. For the selection of a document which has the desired combination of subjects, the respective Uniterm cards must be visually compared. Only the shortest column needs to be considered to find the common number and thus over seventy per cent of the entries can be eliminated from the outset. The posting is tedious as numbers have to be written clearly, but Gull has developed a technique which, with the help of a ticket printer and photography, accelerates the work and eliminates mistakes (38).

For the terminology of the Uniterms, Taube has established a simple rule: "Enter every work in a Uniterm coordinate index system as a filing word on a single Uniterm card. Whenever in a particular system a word is used in one, and only one, descriptive phrase, enter that word as the filing word on a card, followed by the remaining word or words in the phrase. The word or words following the filing word on any card will themselves be filing words on other cards."

The resulting economy in cards as compared with traditional subject headings is considerable. The sub-

ject headings of the Technical Information Division of the Library of Congress were converted into uniterms, and 3620 uniterms plus 720 references for synonyms replaced the former 25,000 subject headings and 24,000 cross references (39).

The center field of an edge punched card has ample space for conventional recording of information. It is also possible to insert a microfilm in this place, a procedure which has been developed by Filmsort (40). A German documentalist proposed to trace a microfilm on a marginal punched card by heliographic methods (Lichtpausverfahren) and suggested that reading could be accomplished with the help of a simple magnifying glass (41).

Frequently the abstract of a given title or document is typed in the center of the card. Many information services use this method to give their subscribers ready access to the literature of the field. The National Association of Corrosion Engineers, for instance, has offered an "abstract punched card service" since 1951 which issues about 2100 cards yearly. Conventional McBee cards, 5" x 8" are used with a double row of holes. The abstract is condensed into about 200 words (42).

An interesting example of the efficient use of the center field for abstracts, charts, drawings, etc. is the punched card catalog of aerodynamic measurements, published by National Luchtvaartlaboratorium in Amsterdam. Marginal punched cards are employed with one row of 168 holes around the edges. Direct coding is used in forty-seven positions to identify aspects which occur frequently: six holes indicate the year of publication and twelve perforations are reserved for the first three letters of the author's name. Nine holes code the sub-group, eighty-four holes are given to superimposed coding of aspects and the remaining nine are kept in reserve. As manual sorting of a large file is cumbersome, the cards are pre-sorted in eleven sub-groups (43).

Various methods are feasible for the duplication of

marginal punched cards (44). The notches have to be reproduced in a separate operation, single cards with the hand punch, larger files with a gang punch. Xerography, a dry photocopying method employed by many libraries for the reproduction of catalog cards, is a frequently recommended method for copying the text. The Ozalid process requires a translucent master which is copied on sensitized paper: the McBee Company furnishes both translucent and sensitized punched cards. A similar photographic process is Copyflex: a translucent master is copied on a punched card sensitized with diazo dyes. These cards, too, are manufactured by the McBee Corporation.

Direct transfer of information from a marginal punched card to an electric typewriter has been developed by Stubenrecht in his "Schreibende Randlochkarte" (writing marginal punched cards) (45).

Marginal punched cards have been applied widely in preference to the conventional card catalog to achieve better and faster bibliographic control. There are hundreds of installations all over the world, most of which report satisfactory results (46). The majority of punched card files is concentrated in the fields of engineering, the exact and applied sciences and medicine; they can be found in numerous libraries especially in the Circulation and Acquisitions Departments, and they are used in many phases of business, for instance warehouse and sales control, personnel work and market research.

The cards are durable (fifty per cent rag content), coding mistakes can be easily mended by "card savers," and the life expectancy is probably better than that of catalog cards. Filing cabinets are offered by most manufacturers of cards, no specially designed drawer is necessary, and 800 cards will fit easily in a fourteen-inch drawer. Measured in terms of space needs punched cards are more economical than the dictionary catalog, but extravagant if compared with a printed index (47). An average of five cards of the conventional size 3" x 5" is needed to bring out all the important aspects of one title, occupying 1.12 cubic inches in

the dictionary catalog; a notched punched card of the size 5" x 8" needs 0.64 cubic inches: five entries in the annual index of the Chemical Abstracts occupy 0.004 cubic inches. However, no conclusion should be hastily drawn from this comparison because the three types of bibliographic control differ in information content and ease of retrieval.

No precise data are available which would enable us to calculate even approximately the budgetary provisions for setting up and using a marginal punched card file. Perry and Kent (48) have worked out the basic mathematical principles of cost analysis, but their thought-provoding research is a report on work in progress and not a "definitive statement of findings." Moreover the authors were interested primarily in the basic theory of information retrieval and their equations and charts are not meant to be translated into dollars and cents. Thorne (49) gave an interesting analysis of efficiency (probability of success) and cost (making and using the file) but he concluded that "cost figures are not representative...and no comparison of the various systems can be made."

A number of the facts involved are not easily measured and the results almost defy generalization. A good example of the difficulties encountered in Marjorie Hyslop's excellent analysis of cost data and subsequent breakdown in four categories:

1. Cost of setting up the system.
2. Cost of equipment.
3. Cost of encoding.
4. Cost of retrieval.

The first group represents the work involved in establishing the classification and determining the terminology of aspects: "the cost is indeterminate but it should not be underestimated." This task demands well trained professional personnel and the expenses will be accordingly high. Miss Hyslop estimates that the preparation of the classification for metallurgy took two years and "the price would have been high in the five figure bracket."

Equipment can be calculated easily. The cost factor is almost negligible and a couple of hundred dollars will suffice to set up a file of 5000 punched cards (51).

The cost of encoding, however, brings us back again to the realm of question marks. "No effort has been made to determine actual figures for this step since they will vary considerably depending on the type and size of the file maintained." The great variety of estimates mentioned in the literature substantiate Miss Hyslop's judgement: as a broad generalization, one could suggest at least half to three-quarters hour professional work and one-quarter hour clerical work per title encoded.

We are on slightly firmer ground in discussing cost of retrieval. Translating the question into the appropriate code is a professional task and may take about five minutes. The sorting of the cards has to be made in batches of 200. The needling time is about one-half minute. However, aligning the cards and returning them to the drawer easily takes one half minute: a file of 5000 cards will, therefore, need about one-half hour sorting time. If more than one needle should be applied, the needling time is correspondingly higher.

Summarizing the above analysis, we must conclude that we have exact cost data for the inexpensive items but only vague conceptions of the expensive ones. Furthermore, we have to acknowledge that the establishment of a punched card file is more expensive than the conventional card catalog. Whereas a title can be cataloged for about $3.00, the preparation of a punched card would be about double that amount per title processed. The reason for this price difference lies in the availability of standardized tools such as classification tables, subject heading lists and printed library catalogs, none of which are applicable to punched cards. Therefore, we cannot expect to decrease our cataloging costs by substituting punched cards for the conventional card catalog.

Cost factors by themselves, however, are meaningless; they have to be considered in conjunction with the

final result, - the successful retrieval of information. Bibliographical control by punched cards has certain advantages: a greater combination of aspects, multiple access points on one card and thus a decrease in the size of the file and finally elimination of precise filing as the cards are not kept in any specific sequence. With the insert of microfilm in the center space of the card, its information content will exceed by far the specifications given by the descriptive cataloging of the conventional card. The combination of notched card and microfilm has far-reaching possibilities.

"A more recent development in the field of marginal punched cards is the use of sheet microfilm bearing ten, or twenty or more frames of text at the usual intermediate reduction ratios, plus marginal notching for sorting. This appears to be coming closer to a complete cycle searching operation, and with higher reduction ratios could store a whole book together with all of the sorting aspects in the form of marginal punches" (52).

The disadvantages of hand manipulated cards lie in the necessary limitation of the size of the card file and the number of aspects. The upper limit is in the neighborhood of 10,000 cards; hand-sorting of larger files would be too time consuming. It is possible to divide a file in sub-groups, but unless these groups are mutually exclusive, the advantages of searching aspect-combinations are lost. Instead of a hand needle which necessitates sorting in batches of about two hundred cards, a simple sorting machine can be employed which permits sorting of up to eight hundred cards (McBee selector: two hundred and fifty cards) with multiple needles in one operation. All these auxiliary methods can increase the quantity of cards but cannot basically change the inherent characteristics of a limitation in size.

The restriction in the number of access points can be partly overcome by superimposed coding. Whereas in direct coding the number of aspects cannot exceed the sum total of perforations, superimposed coding does not have this limitation. Restraint in the number of aspects to be used is necessary, however, lest retriev-

al become so cumbersome that all advantages of the system are lost.

Aspect cards presuppose that the books or documents are arranged by current numbers. This feature limits the application of the method, at least in this country, to a collection of documents, reports, reprints, etc. Aspect cards could be used to control the subject content of storage libraries, but it is doubtful whether the infrequency of use, - an assumed characteristic of books in storage, - would warrant the expense involved. Another disadvantage, at least until Mr. Stern's Microcite is fully developed, is the necessity to consult a shelf list in order to complete the literature search. A further criticism has been that only a small part of the total card space is being used. Theoretically this method has no limitations as to the size of the collection and the number of aspects. A large file of aspect cards, however, would make the continuous refiling cumbersome, unless we add sorting of the file by marginal holes. A large collection of documents would impel us to have multiple cards for each aspect, to increase the size of the card or to acquire expensive reading and punching equipment. As information on the application to larger holdings is lacking, we can only state that the method has been most satisfactory for small and middle-sized collections, (under 1000 aspects; under 25,000 documents).

Hand manipulated punched cards often have been unfavorably compared with machine searching and snobbishly called "the poor man's IBM." It is true that these installations are far less expensive than electronic equipment but that does not imply that they are a cheap substitute. They have been very useful as initial experimental steps to be converted later into a fully mechanized operation (53), but this ancillary rôle does not give full justice to the merits of the system. If management of <u>large</u> masses of material were the sole function of modern methods of bibliographical control, then hand-manipulated punched cards would have to be assigned to a second-rate status. Machine operations have no limitations as to the size of the card file; the sum total of access points on one card, however, is

restricted to the 960 bits on the IBM card. Electronic equipment is very fast, but the breath-taking speed of the machine becomes less impressive if we heed Shaw's repeated warnings "not to confuse part of an operation with the whole operation."

For the control of a limited number of literature references (up to 10,000) hand-manipulated punched cards deserve a leading position. A similar claim can be made for the application to a "closed or one time project which would not warrant the experimentation involved in choosing the most suitable machine system but in which a nearly ideal setup can be devised in almost no time using marginally punched cards" (54).

The decided advantages of hand-manipulated punched cards are at least fourfold:

1. Ample space for conventional recording of information which can be read without transcription and the possibility of inserting a microfilm.
2. Ease of operation. Everybody can quickly learn how to handle the sorting needle.
3. Trifling capital investment and inexpensive current costs.
4. Scant space requirement for the installation.

Aspect cards permit rapid selection and give under certain conditions the most economical operation. The arrangement by current numbers is the most inexpensive shelving system and if the collection is small, direct access to the literature can be given and the necessity of consulting a shelf list would be eliminated.

No system can claim to be the answer to all questions, or in all circumstances, and no prediction is warranted at present, that a given method will soon be obsolete and doomed (55). "A wise literature searcher will, therefore, utilize all facilities at his disposal from old fashioned catalog cards and conventional indexes to the speediest electronic computer. Every method has its place and is justified under the proper set of conditions" (56).

The Gmelin Institute is an example of the efficient and harmonious employment of all methods of bibliographical control, and card catalog, hand-manipulated punched card, aspect cards, and electronic equipment contribute equally to the editing of its famous Handbuch (57). In the vast bibliographical organization of the Library of Congress almost all known methods of information retrieval are employed; (non-mechanical, semi-automatic and fully mechanized); the activities of these installations are coordinated by a Committee on Mechanized Information Retrieval.

Problems for Future Research

The literature on hand-manipulated punched cards is very large and increases yearly at a rapid pace. A considerable number of the contributions are of a high scholarly level and have greatly advanced our understanding. Nevertheless, many of the basic conceptions used are unclear, important parameters have not been substantiated by reliable data, and the sum total of knowledge available is still only a fraction of the information needed. The situation is typical not only of the science of documentation; it is characteristic of all fields of intellectual endeavor and inherent in our conception of scholarship.

The first group of problems suggested for further investigation deals with basic research. Many of the questions in this category will remind librarians of the function of the catalog: Who needs information and what type of information? Furthermore, if it is correct, as it seems to be, that the patrons have to be classified in specific categories, what are the characteristics of each group, expressed in terms of information needed? Are our information systems geared to function in the exceptional cases or are we satisfied to meet the average demand? It has been pointed out during the recent International Conference on Scientific Information, that eighty-nine per cent of the search questions involve three aspects or less. Is this judgment based on a valid statistical analysis of all categories of users? Are the information systems used and the terminology of doc-

umentalists needlessly complicated and would we have satisfactory results by employing less elaborate machines and simpler language?

The second group would treat comparisons between the different systems. What are the standards of comparison? What is the definition of a "good" retrieval system? What are the exact data of speed and depth of a completed retrieval? What is the break-even point between the different methods (58)? What are the characteristic qualities of a given method and for which situation can it claim preference?

The third group would be confined to the specific problems of the hand-manipulated punched cards. What is the optimum size of the card file and the sum total of aspects? How many access points can the single punched card carry without making retrieval unduly cumbersome? How great is the speed of retrieval tested under varying circumstances? What is the optimum space relation between coding area and test? The advantages of the different coding systems in relations to speed and depth of retrieval need better experimental substantiation. More information is needed about the efficiency of multiple rows of perforations.

Little theoretical work has been done on slotted cards and on the combinations of slots and marginal holes. Aspect cards have been tested in small collections but information is needed on their adaptability to large masses of literature.

Conclusion

It is imperative that librarians participate in research on modern methods of bibliographical control. Methods which have been worked out in a non-library situation cannot be adapted without serious disadvantages. "The problems of a library are, for the most part, unique to a library. They should be attacked only by persons who are willing to use them as being unique and to prescribe for them in uniquely suitable terms. There is no reason to think that machines or methods designed

to serve other purposes will be of much direct use to the librarian" (59).

Careful recruitment of qualified personnel, adaptation of library school curricula, and increased opportunities to test the new methods in libraries are essential to stimulate research. Moreover, the profession as a whole must be a receptive audience for the studies and follow with interest all developments in this important segment of librarianship.

*

Equipment

United States

Arizona Tool and Die Company (Boekeler Instrument Company). Tucson, Arizona. Trade name: Needlesort
Edge punched cards with perforations around three edges are manufactured in two standard sizes: 3 1/2" x 6" (48 holes) and 5" x 8" (68 holes). The larger card is available in four colors. The card can also be bought with perforations around the four edges (98 holes). The coding is direct and numerical sequence code.

Burroughs Corporation -- Todd Company Division, formerly Charles R. Hadley Company, Los Angeles, with many regional representatives. Trade name: UniSort
Standard cards available are edge punched cards with one row of holes, 4 holes to one inch, in sizes 3" x 5" up to 6" x 8". For installations which use cards in quantities of 50,000 up special cards are almost the same price as the standard ones. Used frequently for accounting procedures. A "universal" library card has been designed by M. E. Putnam (University of Washington).
The company has been formerly connected with the McBee Corporation and produced the "Rocket" card. This card is no longer manufactured.

Documentation Inc., Washington, D. C.
The firm has developed the Uniterm or Coordinate Indexing System. All installations are tailored according to the specific needs of the given collection and are carefully supervised. The firm is interested in all types of subject cards.

E - Z Sort System, San Francisco, California (60)
Edge punched cards are available in a great variety of sizes up to 8" x 10 1/2" with one to four rows of small holes, 6 to one inch. Holes are staggered along the margin and not in one straight line. All four coding systems are applicable, however E - Z cards are at their very best in direct coding. The multiple row arrangement permits the coding of the greatest number of non-exclusive aspects per inch of edge space of all systems. Available is also a combination E - Z Sort and IBM card.

The system is widely used in research files. The card with a combination of double and triple rows at the Oak Ridge Laboratory mentioned above is E - Z Sort. (21) Other important applications are: The American Society of Metals, Special Library Association, Metallurgical Literature Card, The Paint and Varnish Literature Card, The Illinois E - Z Sort Anaesthesia Record Card, etc.

Frazier Precision Instrument Company, Silver Spring, Md.
Card punch and reader for the Peek-a-Boo system of the National Bureau of Standards.

Gaylord Bros. Inc., Syracuse
Conventional 3"x5" library cards can be furnished with 6 holes punched at top and bottom edge. They can be used for direct coding and numerical sequence code. The cards are applicable for circulation control in small or middle sized libraries.

International Business Machines, Inc., New York
Peek-a-Boo cards with 480 perforations.

Jonker Business Machines, Washington 15. Trade name: Matrex
Peek-a-Boo cards with a capacity of 10,000 or 40,000 positions. A previous model with 15,000 positions is no longer recommended. The firm is also available on a consulting basis for the installations of all types of machine retrievals.

Le Febure Corporation, Cedar Rapids, Iowa. Trade name: X-Ray Sort Card
Cards in varying sizes can be supplied up to 6 1/2" x 10 3/4". Holes are punched in one row, four holes to one inch. Coding is direct, in some instances the sequence code 7 4 2 1 has been employed.

The system is geared for the control of business records.

McBee Corporation, see Royal McBee

Remington Rand, New York 19
Peek-a-Boo card with 640 perforations.

Royal McBee Corporation, Athens, Ohio, with many regional representatives. The Corporation was formed in 1954 with the merger of the Royal Typewriter Company and the McBee Company; McBee cards have been manufactured since 1933. Trade name: Keysort
Marginal punched cards.

Cards are sold in varying sizes from 2" x 3 1/2" to 8" x 10 1/2", also larger cards can be supplied. They are preperforated around the edges with a single or double row of holes. The holes are spaced either on 1/4" centers, or on 2/10" centers, which gives four or five holes respectively to one inch. Recently interior punching has been added and automatic data processing can be achieved with the Keysort Tabulating Punch. The largest manufacturer of marginal punched cards in the United States. McBee cards are so widely used in industries and colleges (both administration and library) that the name has become almost synonymous with edge punched cards.

Superior Business Machines, Inc., New York 17 (61)
Trade name: Flexisort
Does not use preperforated cards. All cards can be used, even existing records can be converted into a marginal punched card file. The Flexisort machine punches the holes and codes by notching in simultaneous operation. 32 holes can be punched on

any one side, (total: 128 holes), five perforations measure one inch. Coding is direct, the alphabetical and numerical sequence code 7 4 2 1 can be employed.

The machine is available on an annual rental basis.

Underwood Corporation, Samas Division, New York
Peek-a-Boo card with 210 and 400 perforations respectively.

William K. Walthers, Milwaukee (62) Trade name: Findex System
Slotted cards. 2 cards are available in an assortment of colors: 6" x 8" and 8" x 8". Special cards are designed for every Findex Installation. (Not manufactured at present.)

Wassell Organization, Inc., Westport, Conn.
Produces a vinyl plastic card 5" x 8" (trade name: Plas-Ta Card) used in the Peek-a-Boo system at the National Bureau of Standards.

Zator Company, Cambridge 38, Mass. (63). Trade name: Zatocoding System
The system employs preperforated cards with forty or seventy-two holes respectively. Both cards measure 5" x 8"; one card carries holes on the top margin, the other one both on top and bottom margins.

The method has many interesting features, but two must be specially emphasized: Superimposed coding based on the random selection of four holes per code position; the use of "retrieval" language for the descriptors instead of "communicative" language. Mr. Mooers, the inventor of the system, does not believe that conventional library classification and subject headings are compatible with successful retrieval.

Selection is made with the Zator Selector.

All installations are tailor-made and carefully supervised by the company.

The equipment is provided on a rental basis.

England

Brisch and Partners, Ltd., London and Toledo (64)
A consulting form for all types of information retrieval and classification.

Carter-Parratt, Ltd., London SW1. Trade name: Brisch-Vistem
Peek-a-Boo card, 6" x 11", capacity: 1000 positions.

Copeland-Chatterson Company, Ltd., London (65)
Trade name: Paramount Punched Card System
Cards are available in a great variety of sizes, mostly with one row, some with a double row of prepunched holes. Two sizes of holes, 4 or 5 to one inch. Coding is direct or combination code. For numerical sequence sorting three code fields can be supplied; a ten position field, 0-9; a six position field, 0-5 and the conventional 7 4 2 1 field. For alphabetical sequence sorting two alphabetical fields are available with 12 or 15 positions respectively.

The largest English manufacturer of marginal punched cards and the pioneer in this field.

France (66)

Compagnie des Fichiers Modernes, Paris 12. Trade name: Rapidtri
Marginal punched card; six sizes are available from 3" x 5" to 7 3/4" x 10 1/2". They have one or two rows of preperforated holes, five holes to one inch.

Direct coding and combination coding are applicable; besides the four position numerical sequence code, there are two alphabetical codes with 18 and 4 positions.

Dequeker S. A., Paris
The fabrication of slotted cards and of the Selecteur has been abandoned for the time being.

Société Détectri, Paris 2. Trade name: Sphinxo
Peek-a-Boo card, 5 1/2" x 10 1/2", capacity: 1000 positions.
Trade name: Statitex
Marginal punched card; a variety of cards is available from 3" x 5" to 8 1/2" x 10 1/2" with preperforated holes ranging from 48 to 134, mostly in one row but some with a double row of perforations.
Trade name: Detectri
Slotted card, available in three sizes from 6 1/2" x 9 1/2" to 8 1/2" x 10 1/2". The coding field is on the right part of the card; the coding positions consist in a horizontal slot.

Société Microdoc, Paris 13. Trade name: Selecto
Peek-a-Boo cards; great variety of sizes and capacities.
2000 punching positions
5000 punching positions, 3 1/4" x 7 1/2", zigzag pattern
8000 positions, 8 1/4" x 6"
12,500 positions, 8 1/4" x 6", on plastic
14,000 positions, 8 1/4" x 6"
20,000 positions, 8 1/4" x 6", on plastic (in preparation)

Société Selection, Vanves, Seine. Trade name: Selectri
Slotted card. Standard size: 3 1/2" x 7 1/2", variations are available, however not exceeding 7 1/2" x 10 1/2".
Coding position consist in a horizontal slot.

Germany (67)

Allform, Berlin W 15
Slotted card. Offered in one size with four capacities, 72, 96, 204, 306 perforations.
The upper part of the card is reserved for conventional recording, the lower part is the coding field. Coding position is a vertical slot.
Peek-a-Boo card; 8 1/4" x 6" with 1860, 2000, 6000 positions respectively.

Edler & Krische, Hannover. Trade name: Ekaha Peek-a-Boo card; 8" x 12" with 7000 punching positions.

Marginal punched card, available in varying sizes with one or two rows of preperforated holes all around the edges. In addition to all conventional coding methods an "additive" code is offered. This method is suggested when selection of a given number is more important than sequence sorting. The coding section consists of three coding fields with eight positions arranged in two rows each. The positions are marked 1 2 4 7 and the fields are designated as in a sequence code with: units, tens, and hundreds.

Slotted cards. Size 8 1/4" x 6", upper part for conventional record, lower part for coding, 210 perforations arranged in eight rows. The coding position is a vertical slot.

Offered also is a combination edge punched and slotted card. The slotted card is the same as described before, but has on the upper edge two rows of holes. Another combination has a single row of holes on the top and halfway down both sides.

Integral, Duesseldorf

Marginal punched card, 8" x 3" with one row of holes around the edges, mainly for bookkeeping.

VEB Organizationsmittel-Verlag, Leipzig

Marginal punched cards.

Available in four sizes ranging from 4 1/4" x 2 3/4" to 11 1/2" x 8" with one or two rows of perforations around the edges. Four holes to one inch.

Slotted cards same formats as above.

Coding field on the lower part of the card, 340 perforations arranged in 10 rows.

Lochkartenwerk Schlitz, Schlitz, Hessen

Marginal punched cards.

17 different varieties available with one or two rows of holes around the edges, four perforations to one inch. Seven colors are offered.

Two combinations of Peek-a-Boo and edge punched card:

a. 48 perforations on two sides and 800 positions
b. 52 perforations on two sides and 2000 positions

Italy (68)

Samo, Milan. Trade name: Selez
Slotted card.
Different sizes from 4" x 2" to 6 3/4" x 5 3/4" with one or two rows of slotting positions on the upper and lower edge of the card. For bookkeeping purposes only.

Japan (69)

Gaikoku Bunken-Sha, Tokyo
Marginal punched cards.

Bunshodo, Tokyo
Marginal punched cards.

Netherlands (70)

Semper Avanti, The Hague. Trade name: Delta Card
Peek-a-Boo card.
The card is 12 1/2" x 9" and has a capacity for 10, 000 positions arranged in a zigzag pattern.

Poland (71)

As far as I could ascertain American type punched cards are used.

Sweden

Esselte, Stockholm. Trade name: Sorto
Marginal punched card.
Different sizes with one row (in a few cases with

a double row) of preperforated holes around the edges. All coding systems applicable, selection by needle. Mostly for business purposes.

U.S.S.R. (72)

As far as I could ascertain Russian documentalists are mainly interested in high speed electronic machines.

Table 1

Summary of Some Typical Industrial and Governmental Mechanical Information Processing and Retrieval Activities from Kent, Allen & James W. Perry - Centralized Information Services. Cleveland, Press of Western Reserve University, 1958. Table 14.

Large Industrial Organization	IBM	Well Satisfied
Large Industrial Organization	IBM	Well Satisfied
Medium Sized Industrial Organization	E-Z Sort Cards	Indeterminate
Medium Size University	IBM	Well Satisfied
Research Institute	IBM	
Dept. of Chemistry, University	Keysort Cards	Well Satisfied
Large Industrial Organization	Keysort Cards	Well Satisfied
Dept. of Biochemistry, University	Keysort Cards	Well Satisfied
Large Industrial Organization	IBM	Well Satisfied
Large Industrial Organization	Keysort Cards	Indeterminate
Large Industrial Organization	IBM	
Government Agency	Peek-a-Boo	Indeterminate
Government Agency	Keysort Cards	Indeterminate
Large Industrial Organization	IBM	Well Satisfied
Large Industrial Organization	Keysort	Partly Satis.
Large Industrial Organization	IBM	Well Satisfied
Municipal Police Dept.	Remington Rand	Well Satisfied
Large Industrial Organization	IBM	Appear Well Satisfied
Government Agency	IBM	Indeterminate
Large Industrial Organization	IBM	Well Satisfied
Large Industrial Organization	IBM	
Large Government Project, University	IBM and Keysort Cards	Well Satisfied with regular file Uniterm. Dissatisfied with experimental Keysort files

Table 2

117 Retrieval Systems

Tabulated From

Kent, Allen, Nonconventional Retrieval Systems in Documentation. Cleveland, School of Library Science, Western Reserve University, 1958

(Air Force Office of Science Research, Technical Note 3)

System	Number of Installations
IBM or similar	57
IBM with Peek-a-Boo	1
Peek-a-Boo	3
Dequeker	4
Zator	1
E-Z	5
Uniterm	11
Uniterm M	1
McBee or similar	34

Table 3

24 Retrieval Systems

Tabulated From

National Science Foundation, Non Conventional
Technical Information Systems in Current Use.

Washington, National Science Foundation, 1958

System	Number of Installations
Zator	1
Uniterm	5
Uniterm M	4
Peek-a-Boo	1
McBee	2
IBM	11

References

1. The author gratefully acknowledges the assistance given by Miss Hickey, research assistant at the Graduate school of library service, Rutgers University, who supplied the initial bibliography.
2. Campbell, D. J. The use of notches in cards as a means of signalling information. Journal of documentation, 9, 1953, 224.
3. The most important references are: Casey, Robert S., J. W. Perry, M. M. Berry and A. Kent, Punched cards, their application to science and industry. 2nd ed. New York, Reinhold, 1958 (hereafter quoted as Casey et al.): International federation for documentation, Manual on document reproduction and selection, The Hague, 1953-1957, 4 volumes (hereafter quoted as F.I.B.): McGaw, Howard F., Marginal punched cards in college and and research libraries, Washington, Scarecrow press, 1952. All three books have extensive bibliographies, the best one is in Casey et al. which lists 276 titles in the first edition of 1951 and continues the numbering from 277 to 677 in the second edition. Another valuable source is Bourne, Charles P., Bibliography on the mechanization of information retrieval, Menlo Park, Stanford research institute, 1957. Excellent summaries are provided by: Pietsch, E. H., Methods of mechanical information, London, 1955 (International congress on documentation of applied chemistry, mimeographed) and Doss, M. P. ed., Information processing equipment, New York, Reinhold, 1955; the most recent one is: Singer, T. E. R., Information and communication practice in industry, New York, Reinhold, 1958. The best bibliography of periodicals is: Brugghen, W. van der, Library and documentation periodicals, The Hague, Inter-

national federation of documentation, 1956. Too late to be utilized for this paper appeared: Perry, J. W. & A. Kent, Tools for machine literature searching, New York, Interscience, 1958 (Library science and documentation I); Taube, Mortimer & H. Wooster, Information storage and retrieval, New York, Columbia University press, 1958.

3a. I am indebted to Mr. Calvin N. Mooers for the list of U. S. patents and to Mr. Roger Connor, Vice President of the Royal McBee Corporation for the information about Mr. Perkins.

4. Perry, James W., A. Kent, M. M. Berry: Machine literature searching, New York, Interscience, 1956.

5. Semper, Gottfried, Der stil in den technischen und tektonischen kuensten oder praktische aesthetik, 2. Aufl. Muenchen, Bruckmann, 1878/9.

6. Bloomfield, M. Evaluation of coordinate indexing at the Naval Ordnance Test Station. American documentation, 8, 1957, 22.

7. Haykin, D. J. What is needed by way of special classification? American documentation, 8, 1957, 51.

8. Stroem, Ivar. On multidimensional classification systems. Retrieval by means of marginal punched cards. U.D.K. bulletin, (Oslo) 3, 1957, 33.

9. Ruston, W. R. Une application combinée des cartes perforées et de la classification decimale universelle pour la bibliographie. Review of documentation, 18, 1951, 31.

10. Grobe, Gustav. Internationale dezimalklassifikation und randlochkarte. Nachrichten fuer dokumentation 5, 1954, 141.

11. Weigelin, E. and J. Ossendorff. Randlochkarte und dezimalklassifikation. Nachrichten fuer dokumentation, 9, 1958, 25.

12. F.I.D. 724.3 E3. See also the critical approach in Casey et al. p. 378 and p. 528.

13. Cherry, Colin. On human communication. New York, John Wiley & Sons, 1957.

Quastler, Henry. A primer on information theory in Yockey, Hubert P. ed. Symposium on information theory in biology, Gatlinburg, 1956. New York, Pergamon press, 1958.

14. See, however, the French systems of slotted cards.
15. Besides the titles quoted in note 3 most helpful were: Peakes, Gilbert L., A. Kent, J. W. Perry, Progress report in chemical literature retrieval. New York, Interscience, 1957.
16. Casey et al. p. 438 gives an excellent mathematical analysis of coding systems by Carl S. Wise.
17. Cox, G. J., C. F. Bailey, R. S. Casey. Punched cards for a chemical bibliography. Chemical engineering news, 23, 1945, 1623, also reproduced in FID, 744.23 E1.
18. Mooers, Calvin N. Zatocoding for punched cards. Boston, Zator Company, 1950 (Zator technical bulletin 30) see also FID 744.24 F1.
19. Gilbert, Paul T. An optimum punch card code for general files. American documentation, 9, 1958, 84.
20. McGaw, op. cit. p. 71.
21. Hood, S. L., R. A. Monroe, W. J. Visek. Edge-punched cards for scientific literature references. Oak Ridge, Technical information service, 1953 (U. S. Atomic Energy Commission, ORO-102).
22. Hardy, F. E. M. The capacity of edge punched cards for information storage. Office magazine, 5, 1958, 452.
23. Crosz, Gabor. Spezielle ausfallerscheinungen der mehrreihigen lochkarten. Dokumentation, fachbibliothek, werkbuecherei, 6, 1958, 117.
24. Wise op. cit. in Casey et al. p. 462.
25. FID 744.6
26. FID 744.4; Casey et al. p. 125.
27. Holmstrom, J. Edwin. Facts, files and action. London, Chapman & Hall, 1953.
28. Wildhack, W. A. & Joshua Stern. The peek-a-boo system-optical coincidence subject cards in information searching (in Casey et al. p. 125), the most authoritive paper on the subject written so far.
29. U. S. patent 1, 165, 465 (Dec. 28, 1915); Kistermann, Friedhelm, Zur geschichte und entwicklung des sichtlochkartenverfahrens, Dokumentation, fachbibliothek, werkbuecherei, 6, 1957, 7.
30. First described on the 22nd ASLIB conference, London, 1947, 37, see also Batten, W. E., Spec-

ialized files for patent searching (Casey et al. First edition, 1951, 169).

31. Loosjes, Th. P. The Delta card. ASLIB proceedings, 9, 1957, 142, also printed in Nachrichten fuer dokumentation, 8, 1957, 42. See also Condonnier's reply in ASLIB proceedings, 9, 1957, 142.
32. Wildhack, William A., Joshua Stern, and Julian Smith. Documentation in instrumentation, American documentation, 5, 1954, 223, reprinted in Taube, Mortimer and associates, Studies in coordinate indexing, 3, 1956, 114.
33. Casey et al. p. 139. Another method for cancellation is proposed by Carl Braband using a cancellation card of green paper. Carl Braband, Einige vorschlaege fuer die ausgestaltung des sichtlochkartenverfahrens, Nachrichten fuer dokumentation, 8, 1957, 42.
34. Gagarin, R. Die sphinxo-sichtlochkartei fuer ein teilgebiet der anorganischen chemie. Nachrichten fuer dokumentation, 8, 1957, 148.
35. Rothschuh, K. E. Eine sichtlochkartenmethode zur erfassung einer grossen sonderdrucksammlung. Nachrichten fuer dokumentation, 9, 1958, 90.
36. Microcite an aid to more efficient referencing. National bureau of standards, Technical news bulletin, 41, 1957, 141, see also Journal of documentation, 14, 1958, 40.
37. Taube, Mortimer and associates. Studies in coordinate indexing. Washington, Documentation Inc., 1953-57, 4 volumes; see also Casey et al., p. 153.
38. Gull, C. D. Posting for the uniterm system of coordinate indexing. American documentation, 7, 1956, 9.
39. Thomas, R. B. and C. D. Gull. The choice of terms for a uniterm coordinate index of scientific and technical reports. In Taube and et al. op. cit. 1, 1953, 47.
40. Casey et al. p. 74.
41. Thiele, Georg. Rationelle uebertragung von schrifttum auf randlochkarten durch das lichtpausverfahren. Nachrichten fuer dokumentation, 7, 1956, 216.
42. National association of corrosion engineers, NACE corrosion abstract punch card service, Houston,

The association, 1958.
43. Kock, A. C. de. Description of the N. L. L. card catalogue of aerodynamic measurements. Amsterdam, Nationaal Luchtvaart-laboratorium, 1955 (Report F. 173); Kock, A. C. de and A. C. F. Hagsdorn, The N. L. L. card catalogue of aerodynamic measurements, Manual, Amsterdam, Nationaal Luchtvaartlaboratorium, 1956 (Report F. 178); See also short description by R. A. Fairthorne, Journal of documentation, 10, 1954, 11.
44. Gull, C. D. Transcription problems in preparing and using punched card files. Casey et al. p. 555.
45. Stubenrecht, Alfred. Die Bedeutung der technischen information fuer die produktivitaet des betriebes. Nachrichten fuer dokumentation, 9, 1958, 70.
46. Harksteen, Barbara L. Review of applications. Casey et al. p. 303.
47. Shaw, Ralph R. Management, machines, and the bibliographical problems of the twentieth century in Shera, Jesse H. & M. E. Egan, Bibliographic organization, Chicago, University of Chicago press, 1951, p. 200.
48. Perry, J. W. & A. Kent. Documentation and information retrieval - An introduction to basic principles and cost analysis. New York, Interscience, 1957.
49. Thorne, R. G. The efficiency of subject catalogues and the cost of information searches. Journal of documentation, 11, 1955, 130.
50. Hyslop, Marjorie. An edge-notched punched card system for metallurgy in Shera, J. H., A. Kent, J. W. Perry, Information systems in documentation. New York, Interscience, 1957.
51. Ohlman, Herbert. Low cost production of marginal punched cards on accounting machines. American documentation, 8, 1957, 123.
52. Shaw, Ralph R. Mechanical storage, handling, retrieval and supply of information. Libri, 8, 1958, 1.
53. Cirkler, F. & E. H. Hungermann. Patentdokumentation mit handlochkarten. Nachrichten fuer dokumentation, 9, 1958, 123.
54. Personal letter of Curtis L. Brown, editor of the

Bulletin of the Institute of paper chemistry, to the author.
55. Miller, Stanley C. Punched cards are dying. Machine accounting and data processing, 1958, 15.
56. Brown, Curtis L. Two approaches to the retrieval of information from a special library. In Peakes, Gilbert L., A. Kent, J. W. Perry, Progress report in chemical literature retrieval, New York, Interscience, 1957.
57. Pietsch, E. H. E. Evaluation of mechanized documentation at the Gmelin Institute in Casey et al., p. 571.
58. The efficiency of indexing according to various systems is investigated by a current ASLIB project. See National science foundation, Current research and development in scientific documentation, no. 2, Washington, National science foundation, 1958, 1.
59. Ridenour, Louis N., R. R. Shaw, A. G. Hill. Bibliography in an age of science. Urbana, University of Illinois press, 1951.
60. F.I.D. 744.25; Casey et al., p. 100 & 161.
61. F.I.D. 744.26
62. F.I.D. 744.64
63. F.I.D. 744.24; Casey et al., p. 340.
64. Kyle, Barbara, E. G. Brisch. Something new in classification. Special libraries, 47, 1956, 100.
65. Ashworth, Wilfred, ed. Handbook of special librarianship and information work. London, Aslib, 1955, p. 310.
66. Salvan, Paule & Paul Poindron; Les systemes de selection, Bulletin des bibliotheques de France, 2, 1957, 455.
Centre d'information du material et des articles de bureau, Appareils et fiches perforées pour selection manuelle. Paris, C.I.M.A.B., (1956).
In January, 1959, Microdoc will publish a pamphlet: La 'selection' et la méthode 'selecto' (Procédés cordonnier).
67. Draheim, Heinz. Handlochkarten im liegenschaftskataster, Berlin, Wichmann, (1956) reprint.
Kommunale gemeinschaftestelle, Die handlochkartei in der kommunalverwaltung. Koeln, 1957.
Lingenberg, Walter. Ueber die anwendung von lochkartenverfahren in bibliotheken, Koeln, Greven-

verlag, 1955 (Arbeiten aus dem bibliothekar-Lehrinstitut des Landes Nordrhein-Westfalen, 9) Scheele, M. Die lochkartenverfahren in forschung und dokumentation mit besonderer beruecksichtigung der biologie. Stuttgart, Schweizerbart, 1954; Rev. Journal of documentation, 1955, 89. Stegmann, Anton. Die verwendung des lochkartenverfahrens bei der flurbereinigung. Stuttgart, Ulmer, 1957 (Schriftenreihe fuer flurbereinigung 12). I was not able to utilize: Ausschuss fuer wirtschaftliche verwaltung, Die handlochkarte. Technik und anwendung in wirtschaft and verwaltung, Frankfurt, Agenor Verlag, 1958.

68. International conference on documentation, 18th, 1951, La documentazione in Italia, Roma, Consiglio nazionale delle ricerche, 1952.
69. Hirayama K., M. Masuyama and S. Nakumara ed. Panti-Kado no riron to jissai (Theory and application of hand sorted punched cards). Tokyo, 1957.
70. Pietsch, Erich, G. Hantke, G. Mulert. Massgebliche dokumentationsein-richtungen in den Niederlanden, Nachrichten fuer dokumentation, 5, 1954, 13.
71. Majewski, Zygmunt. Der stand der dokumentationsarbeiten in Polen. Nachrichten fuer dokumentation, 9, 1958, 15.
72. Cherenin, V. P. & B. M. Rakov. Experimental information machine of the Institute of scientific information of the USSR Academy of sciences. Moscow, 1955.
    Rakov, B. M. & V. P. Cherenin. Machines for retrieving information in the USSR, Unesco bulletin for libraries, 11, 1957, 192.

Volume Four   Part Two

FEATURE CARDS
(Peek-a-Boo Cards)

by

Lawrence S. Thompson

1. Definition

1. Some American documentalists have used the rather ludicrous terms of "peek-a-boo" (61), "peekable," or "peephole" for the type of punched card to be discussed in this essay. Other English-speaking documentalists have played with such other terms as "super-imposable," "optical," or "coincidentally" punched cards. Foreign terminology is less uncertain: "fiches superposables" (Fr.), "sichtlochkarten" (Ger.), "titthålkoft" (Sw.), or "onderwerpponskaarten" (Du.). They are also called Cordonnier or Batten cards after two of their leading exponents. The term "feature card" is advocated by J. L. Jolley (34) and J. Edwin Holmstrom (correspondence with the writer). Contrary to other punched card systems that use one card for one document or one item, the feature card system uses one card for one subject, characteristic, aspect, or feature. Hence the more exact (and more dignified) name will be used here.

2. Horace Taylor's Patent

2. The earliest record of the use of feature cards is 1915, when Horace Taylor of Brookline, Mass., patented a "selective device" for the identification of birds (56). He used a foundation sheet bearing the names of birds in combination with "screen sheets" showing such characteristics as having topknot, perching on a branch, medium in size, and blue or blue-gray. Each screen sheet is perforated over the positions on base sheet if the bird on the base sheet has these characteristics. When the four screen sheets thus perforated are placed over the base sheet, the name of the blue-jay only may be read, since it is the only bird combining all of these characteristics. Although Taylor used a base sheet rather than coding his perforations, he hit upon the basic principle of feature cards that has not been changed.

Four decades later there was a return to the notion of a base sheet in greatly refined fashion, microcite (47), infra.

While the illustration on Taylor's original patent showed base sheets of only thirty-five birds he stated that the perforations on each of the screen sheets were "so arranged" that the screen sheet could be superposed on the foundation sheet in four positions, that is, reversible top for bottom, or recto for verso (with both top and bottom aspects on the verso). This arrangement increases the potential of Taylor's "selective device" fourfold. Taylor does not describe any equipment for making his rectangular perforations.

Taylor does not describe the arrangement of his screen sheets. While classification or other arrangement of the feature cards is, obviously, an essential to their effective manipulation, (8; cf. also 4), the theoretical aspects of the problems of classification belong in another essay and will be treated here only as they may affect mechanical aspects of searching.

3. H. E. Soper's Patent

3. In 1920 Herbert Edward Soper of London, England, patented another feature card system (52). He does not suggest any specific use in his patent other than for "tabular and statistical data" in general. He does not use a base sheet. His card, as illustrated in the patent, shows only 130 positions but there is no implication that the size of the card is limited to this number. He uses simple circular perforations made by an awl-like perforator. He suggests a luminous screen at the back of the superposed cards, a readout device still widely used. He says categorically that his device "possesses great advantages...over systems of indexing employing cross references or multiple entry." He emphasizes the savings in cost and material. Soper's invention is the classic statement of the feature card in generalized form. It is curious to note that neither Soper nor Taylor seem to have had any subsequent influence, and they have only been noted historically in

the accelerated interest in feature cards that has developed in the post-World-War II era (e.g., 31, 60).

4. The Borgeaud-Liber Patent

4. A third unrelated patent, taken out in France by Borgeaud et cie. and Henry Liber (6) in 1924 did bear fruit in the work of Cordonnier some two decades later (3, 30), infra. The Borgeaud-Liber patent is quite similar to Soper's, and it uses examples of identification of individuals. Although Soper did not imply that his cards were positively limited to 130 positions, it is significant to note that the Borgeaud-Liber cards show 1,000 positions. This patent shows a new refinement in the use of different colored cards for chronological identification. He cites the example of a record of orders by a business firm in which a given individual may be identified as a client, while the position to which he is assigned is backed up not by a transparent perforation, but by a color showing the penultimate month to the month of his last order. Holmstrom (22) expressed this use of color in more general terms. If we need to ascertain documents which satisfy criteria A, B, and C, but not Z, we need simply to put Z at the bottom of the superimposed pile and cover it with a transparent colored plastic (cf. also 7). Another use of color, of course, is to differentiate various series of references after the maximum number of positions on one series of cards is used (4, 22, 27, 43).

5. Feature Cards for Mineral Identification

5. In August 1920 C. J. Gray, a geologist in the service of the Zululand and Natal Mines Departments, read a paper to the Geological Society of South Africa on the use of physical characteristics of minerals for their identification (25). There is no evidence that he was acquainted with the work of Taylor and Soper, and for the next quarter of a century mineralogists operated independently, insofar as their publications (16, 17, 18, 21, 28) indicate, of the ideas of the three early patentees. The problem faced by mineralogists is that some

rare mineral, or some deviant or unusual form of a common mineral, may resemble a sample. This possibility may be ignored in confirmatory tests. Gray's experience with tables of characteristics for identification of minerals had been unsatisfactory, since they do not invariably allow the scheme of examination to be adapted to a particular specimen.

Gray set up a list of 361 minerals and presumably had only that number of spaces for perforation, or perhaps a few more. He selected 66 characteristics, such as luster, color, cohesion, hardness, effect of heat, specific gravity, and crystallization, and subdivisions under each. Like Taylor, he used a base sheet. He was cautious in his claims for his system: "No claim is made that every mineral can be determined exactly by use of the set of sheets alone, but such use will rapidly so reduce possibilities that almost invariably the exact identification will be clear to a man with a good knowledge of minerals..."

Gray had no more direct influence than Taylor or Soper; and when J.D.H. Donnay, a Belgian mineralogist, first explained his system sixteen years later (17), he seems to have been ignorant of Gray's work. Donnay, who worked out his system with J. Mélon of Liege, operated with cards with 330 positions (representing some 360 species) in a fashion almost identical with Gray's. He is more positive than Gray about the accuracy of his system. He argues that if two or three positions are open at the end of the operation, the selection of supplementary properties will eliminate all but one. In a later article for American mineralogists (16) he was more conservative and warned that it was hard to say that a mineral has no cleavage from examining one specimen only.

Donnay made his cards from Manila cardboard and filed them in a wooden box with an obliquely cut lid. Donnay and Mélon also made a larger set, using 210 cards for characteristics, but it seems to have been too expensive to produce in quantity (18).

Hurlbut (28) developed a system of mineral ident-

ification on edge-notched cards, with each card representing a mineral rather than a property of a mineral. He said that it was easier to separate by properties according to his system than to thumb through the list of properties represented by each of Donnay's feature cards. He also pointed out that his cards could be thrown together in any order after use. Wachtel (57, 58), infra, rejected edge-notched cards for identifying properties of nuclides, since the total bulk of the cards would be twice as great if she had to provide one card for each document. The same is true of Hurlbut's system, except that he would have approximately six times as many cards in his pack as Donnay had in his, if there must be a card for each of 361 minerals.

Fairbanks (21) developed a somewhat more elaborate system than Donnay's for the identification of non-opaque minerals, especially the fine intermixed and disseminated forms of ores. He used standard tabulating machine type cards (Powers 1060, 7 3/8" x 3 1/4", purchased from Remington-Rand), but he assigned only 356 of the 540 positions to known mineral species. There are 117 cards in a set. Like Donnay, Fairbanks argues for the superiority of the feature cards over conventional tables.

6. W. E. Batten's Feature Cards

6. The work of the mineralogists with feature cards attracted little, if any, attention outside of their own field. The present enthusiasm for feature cards may be attributed mainly to the work of W. E. Batten (3, 4), the head of the Intelligence Department of the Plastics Division of Imperial Chemical Industries, London, and G. Cordonnier, a former professor of mathematics and chief engineer for the Génie Maritime, Paris (8). Résumés of the work of Batten and Cordonnier, with notes on the later development of their work, may be found in Holmstrom's articles in the F.I.D. Manual on Document Reproduction (22), in his articles in the UNESCO Monthly Bulletin on Scientific Document Reproduction and Terminology (mainly 54), and in the second part of his work on Facts, Files, and

Action in Business and Public Affairs (27).

Batten came face to face with the problem of indexing patent literature pertinent to his field in 1939 when the British Patent Office ceased publishing its Official Abridgements. Due to the relative inflexibility and expense of high-speed mechanical devices for sequential handling of large collections of data, Batten had to turn to some other device to ascertain patents which dealt with specific topics. He first developed feature cards providing spaces for 400 documents, numbered 1 to 400. When this series was complete, he used a second series with positions numbered 401 to 800, and so on until he could provide for 4,000 items, the tenth card in the series providing for positions 3,601 to 4,000. By this time the basic limitation of the primitive feature card, namely, the number of documents that could be searched in one operation, became apparent. Batten then substituted for his eleventh and subsequent series of documents a standard Hollerith (IBM) card with 800 positions. (Note that these cards could not be used for machine selection with Batten's notation, that they were only a handy, readily available medium for a feature card system.)

Batten stated categorically that "No amount of mechanical aid will make up for a defective classification system." He made a genealogical breakdown of the subject matter into classes of aspects and subclasses. Then he used a suitable decimal classification so that each major or minor aspect class was coded with a sequence of digits. Batten's rather classical treatment of his subject matter illustrates a specific problem that every user of feature cards must face and solve.

Holmstrom (22) suggests various forms of headings for a subject classification of feature cards. One suggestive point is that the "facets" in Ranganathan's Colon Classification might lend themselves to this treatment.

Batten and others recognized serious limitations to his system. Batten pointed out that the speed of searching is slow, less than in a fully mechanized system, and he stated that his system is "essentially for the small-

scale operator" (3). He said that it saves technical man-hours but not clerical man-hours. A graver limitation is that only so many documents can be recorded, a more aggravated problem with him than with Cordonnier and others, since we will note cards providing for up to 40,000 documents. Feature cards have been contrasted with another widely used manual punched card system, edge-notched cards (27, 54; cf. also paragraph 5, supra). While the number of documents that can be indexed by the latter is theoretically unlimited, the number of subjects is limited by the number of notches that can be placed on the edges. In the case of the former, the number of subjects to be indexed is theoretically unlimited (Batten used about 1,000 cards per series [22, 27]), the number of documents that can be indexed in a single series is limited to the number of punchable positions on the feature card.

Batten (3) did not think his system could be used with a large number of references. Nevertheless, he had some vision of wider application, suggesting that a different mechanical form of the same principle would be the solution. For example, it might be possible to replace feature cards with feature strips, "the latter consisting of perforated ribbons wound on spools and adapted to be run through a scanning device whilst superimposed." Or the signalling medium need not be a perforation on a ribbon, but a mark on a film or an impulse on a sound track.

7. G. Cordonnier and Developments in France

7. G. Cordonnier gives credit to the Borgeaud-Liber patent in his fundamental study of feature cards (8), stating that Sphinxo cards produced by Détectri, 68 rue de Richelieu, Paris (53), incorporated the possibilities in the patent after it lapsed into the public domain. While Taylor, Batten, and the mineralogists were all attacking a specific problem of selection, the Borgeaud-Liber patent (like the Soper patent) is couched in more generalized terms, and Cordonnier's uses for feature cards are broader than any previous practical

application. Cordonnier first developed the Sphinxo card (8, 53) with 1,000 positions. Although Cordonnier (8) expressed dissatisfaction with the limited number of positions, this card still has practical uses in instances in which the cases are usually limited to less than 1,000 (e.g., medical records, industrial accidents, and even limited bibliographies; see 53). These cards are available in seven colors, and a mechanical perforator and reading frame (for illuminating the "through holes") are available. Détectri's promotional brochure (53) is especially persuasive in its description of non-documentary applications of Sphinxo cards.

For most documentary purposes 1,000 positions are insufficient, and therefore Cordonnier developed the Sélecto card (produced by Société Microdoc, 9 rue Rubens, Paris 13^{e}) with 2,000 positions and a perforator to go with it (8, 49, 50). However, it was necessary to cover much larger collections of documents, and the next step was to expand the Sélecto cards to include 12,500 positions. Moreover, this new type of card allowed the subsequent introduction of new entries, for positions could be left to provide for this contingency (22). Cordonnier worked out a card 15 x 21 cm. (5 1/8" x 8 1/4") printed like graph paper (with positions on x from 0 to 99, on y from 0 to 124). They were printed in eight different colors so that a total range of 1000,000 documents could be covered in eight series (27). A special perforator, Per Sélecto, and a reading frame, Sta Sélecto, are available from Microdoc. The small size of the positions for perforations were a source of some concern, since it was feared that shrinkage and expansion might vitiate the fine dimensions of the 12,500 position card. Cordonnier first experimented with plastic sheets, but later he found a special cellulose material which is reported to be satisfactory (27, 50).

Holmstrom (27) reports six different applications of Cordonnier's system in Paris. Cordonnier himself used it at the Naval Ministry as the basis for a technical service. The Mineralogical Laboratory at the Sorbonne used Cordonnier's system for crystallographic identification; and Holmstrom emphasizes the economies

in space when feature cards are used to index references to organic chemical compounds in general (one card for each kind of atom in the molecular diagram instead of thousands for various substances). The judicial identification system has put over 1, 500, 000 fingerprints and other criminal data on feature cards, and searches are now said to be much more rapid than formerly. The Centre National des Recherches Scientifiques indexes and searches its files of unpublished translations with feature cards. The Union Française des Organismes de Documentation uses feature cards as a means of ascertaining which of several hundred libraries and documentation centers in France offer such services as microfilming, publishing abstracts, translating, etc.

The most striking use of Cordonnier's system is by the Institut des Fruits et Agrumes Coloniaux, 6 rue de Général Clergerie, Paris 16^{e} (9, 22, 27, 29, 30). Originally the Institute used the 2, 000 position cards, but it soon changed to the 12, 500 position cards. The Institute indexes 350 current periodicals, using a special type of subject classification developed by Cordonnier (8, 39). At the end of each year, or when all positions on an aspect card have been used, the Paris office punches out several sets. Thus the colonial branches are able to conduct their own literature searches and request microfilm of needed articles from the Paris office. Punching several sets is said to be quick and cheap. Holmstrom (22) says it saves the expense and delay of compiling and printing indexes to abstract bulletins. Further, he says, it might make possible the consolidation of references in several different abstracting organs or bibliographies.

Another biological application of feature cards was developed by Holmstrom for the Fisheries Biology Branch of the Food and Agriculture Organization (26). The original report with details of the recommendation was not available.

Holmstrom (22) records certain apparent objections to Cordonnier's system. While the number of documents that can be searched in a file is limited, this is compensated by the extreme quickness of the search

and the fact that in a large organization, different sets can be searched simultaneously at different desks. Lining up pinpoint perforations (0.7 mm in diameter and 1.4 mm apart) is said to be time-consuming and possibly lead to error; but the manufacture of the cards (Cordonnier's are cut to 0.001 mm precision under controlled temperature and humidity), the accuracy of punching devices and readout devices, and, above all, practical working experience over several years, indicate that the system is quick and accurate (8, 50). Holmstrom thinks there is no difficulty involved in developing a systematic filing system, since cards can be filed in an arbitrarily serially numbered order with filing numbers written in on the margins of an outline of the classification, or any desired classified order; or verbal descriptive headings may be used. They can also be edge-notched for purposes of arrangement, a suggestion also made by Kistermann and Ühlein (37). There is the objection that it is necessary to go to another file to find even the title, but Holmstrom proposes that a file of microfacsimiles of indexed articles be kept within reach. Microcite (47), infra, is another possible answer. It is said that it is hard to search completely for any general subject, but Holmstrom thinks that the converse is true, since each document can be punched on the feature card or cards representing its most specific subject, or subjects.

One grave obstacle is the high labor cost for indexing a large number of articles and the necessity of using meticulous workers (22). Two persons working in tandem cannot punch more than 1,500 holes in a working day. For as many as 3,000 punchings two teams of two workers must be used, one working with odd-numbered feature cards, the other with even-numbered ones.

It should be noted that Microdoc also produces 8,000 and 14,000-position cards, 10,000 and 20,000-position plastic sheets, and 5,000-position cards in standard IBM format which can be filed in IBM equipment (54). The latter has a space-saving rhomboid pattern of positions similar to the Delta cards produced by Semper Avanti (10, 43, 44), infra. Microdoc fur-

nishes books with thumb indexing for filing the cards. There is a hand punch for the 8,000 position cards, but for the others there is a specially designed punch with a screw adjustment in two dimensions to enable quick and accurate location of very small holes (54).

No mechanical problem has been reported in the use of cards with very large capacity when a precision punch is used; and, clearly, they increase the number of items processed by a single operation. However, they do not allow the identification of required items by negative characteristics (22, 7, and supra, paragraph 4), since reflected color will not show through them.

8. The Netherlands

8. The feature card has been used extensively in the Netherlands, and some new advances have been made there. In industry it is used by the Patent Division of Philips' Gloeilampenfabrieken, Eindhoven (59), Hollandsche Signaal Apparatenfabriek (special cards; details unavailable), and Allgemeene Kunstzijde Unie, Arnhem (Delta cards, infra). Th. P. Loosjes of the Centrum Voor Landbouwdocumentatie, Gen. Foulkesweg la, Wageningen, one of the foremost theorists on the use of punched cards in documentation, reports the use of Delta cards in the Institute for Land and Water Management Research and in agricultural field experiments in Wageningen. In a private communication he says he has used Sphinxo cards but he emphasizes their limitation to 1,000 positions.

J. Westendorp (59) applies feature cards to patent literature at Philips in much the same way that Batten does. He uses standard IBM (or Hollerith) cards with 800 positions as feature cards, although he points out that two extra rows on these cards actually allow for 960 punchable positions. Since Philips also uses machine-selected IBM cards, Westendorp is in a favorable position to contrast the manual and the mechanical systems or simultaneous versus sequential scanning over and above the matter of expense of initial outlay for equipment. He is particularly impressed by the possib-

ility of adding as many feature cards as are needed to the pack, whereas IBM and other item cards place definite limits on the addition of subjects to the code. On the other hand, Westendorp uses his machine-selected IBM cards for searching for several ideas at the same time, something that the feature cards do not permit in their present stage of development. Westendorp states that it is his experience that IBM cards wear out under machine sorting more quickly than do feature cards. Finally, Westendorp points out the inevitable disadvantages of feature cards, that a new series must be started when all positions on one are filled, whereas one can add item cards indefinitely and in a single series to a machine selected IBM system.

Delta cards are produced by Semper Avanti, Losduinenweg 507, The Hague (10). The great advantage of the Delta card is that it provides more punching points (10, 000) on the area provided than other feature cards of comparable size except those which use pinpoint perforations (e.g., the U. S. Office of Basic Instrumentation). This is made possible 1. by the rhomboid design of the squares and 2. by the arrangement to punch not in the preprinted areas but on the intersecting lines. The punching machine has a directive needle to facilitate this work, but the punching device is hardly so refined as to be called a precision instrument. The reading of the card is made easier by dividing the card into 100 compartments (with 100 intersections in each compartment) instead of 100 rows on both x and y, printing the fifth line (both horizontal and vertical) in heavier ruling, placing dots at intersections of the third and eighth vertical lines, and the placement of odd vertical lines at a slightly higher position than the even ones. The card measures 32 cm x 23. 5 cm. For these reasons Loosjes (43, 44, 45) argues for the superiority of the Delta card over other feature cards. Loosjes advocates the use of Synoptic filing (cf. 41, 51) as a convenient device for aiding in maintenance of arrangement of feature cards (45), since the signals in this system allow refiling without studying the headings. He thinks the use of Uniterm is desirable for the headings of the feature cards, since it enables the searcher readily to select the cards he wants merely by noting

what numbers they have in common (45). The punched positions could correspond to the numbers written out on the Uniterm card. Thus the possibility of error in visual comparison of numbers on Uniterm cards is eliminated, and it is unnecessary to compare Uniterm cards in several steps and to record all the common members on a blank card at each step.

Loosjes has a deeply rooted faith in the utility of feature cards; and, aside from their technical virtues, he places special emphasis on their ability to reflect the dynamics of scientific research (43, 45). As research uncovers new ideas and new viewpoints, appropriate feature cards may be added to a set; but in the edge-notched cards, the number of subjects that may be handled is limited unless one wishes to work with multiple sets.

9. Sweden

9. In Scandinavia Carl Björkbom has reviewed briefly the literature of feature cards (5), and the Systems Division of Esselte (Bryggargatan 17, Stockholm) has developed and exploited them under the name of "Findex" (23, 24). There are two Findex cards, one with 7,000 positions, the other with 3,000 positions. The Swedish cards offer no different ideas or new applications. The main inspiration comes from the Cordonnier cards (Sélecto), and the firm's organ even contains an article citing an example obviously taken from the experience of the Institut des Fruits et Agrumes Coloniaux (24). The cards show some features of the Delta cards, with heavy guide lines and with division into squares of a hundred positions in the 3,000 position card. However, the cards are punched in the squares rather than at the intersections of the lines. Some of the German feature cards (e.g., Ekaha and Allform, infra) are also said to be used in Scandinavia, but specific information on their applications is not available.

10. England

10. In England and the United States some definite practical advances and, in the latter country, imaginative applications of feature cards have been made. It is of some significance to note that recorded uses of feature cards in the United States are by federal government or federally subsidized agencies. In England, the Netherlands, Sweden, and West Germany, they have been developed commercially.

In England feature cards are known to have been produced or promoted by three firms: 1. Carter-Parratt, Ltd., Iddesleigh House, Caxton Street, London, S.W. 1, produces the Brisch-Vistem card, 15 x 28 cm., with 1,000 positions, "like Ekaha [infra] but smaller cards, with good layout facilitating quick location of positions" (54); 2. Industrial Studies and Investigations, formerly located at 40 A High Street, Hampstead, London, N.W. 8 (27), but present address unknown, and particulars on their cards lacking; and 3. J. L. Jolley and Partners, Ltd., New Road, Great Missenden, Bucks, with an American office under the name of Brisch, Inc., 1070 Union Commerce Bldg., Cleveland. Jolley and Partners are successors to a British firm known as Brisch Indexing, Ltd.; and Carter-Parratt makes and markets the Brisch-Vistem cards.

The Carter-Parratt cards range from 1,000 to 2,500 positions, although this firm is willing to make cards with larger capacity and smaller holes, making allowance for the disadvantages of the larger cards (communication of J. L. Jolley in writer's file). At present, the Jolley-Brisch group puts strong emphasis on the use of the Carter-Parratt cards for personnel records (7a, 34), but there is also attention to other possible applications (34, 36) and to the theoretical structure of feature card systems (35, 36). The Carter-Parratt feature cards are recommended for special libraries (7a, 33) and for operational research, medical research (cf. 1, 11, 12, 13, 14, 15 and discussion, infra), hospital records, market research, photographic print and negative indexing, social surveys, property records, fault recording and correlation, and criminal

records (7a).

The literature available on Carter-Parratt cards and their applications is exclusively from the manufacturers and promoters and is thus based on successful applications.

The following claims are made in the promotional folder (7a):

> High speed of operation. Information is obtained within minutes. For example, all personnel possessing (say) four required characteristics can be selected from 1,000 individuals in about ten seconds.
>
> Constant control of data. The user has absolute control over the data at all times because the unique sorting process requires the cards to be absent from their place in the record for a few seconds only.
>
> Simplicity. After initial punching, all possible sums and correlations are immediately available.
>
> Economy. The initial cost is reasonable, and maintenance costs are negligible. The characteristic cards can be used almost indefinitely.
>
> Compactness. Up to 1,000 elements of information can be punched on one characteristics card, and 1,586 such cards can be filed visibly in a unit 33" x 24".
>
> Versatility. The system caters for both changing and static data and can therefore be applied to a wide variety of records.

This summary of the virtues of feature cards clearly applies to pertinent situations that the promoters have found in their work with specialized classifications.

Jolley has made certain comparisons of the feature card with the item card which are rather unfavorable to

the latter (33). He re-emphasizes the basic fault of a system involving item cards, that it is impossible to add a new feature after the feature field is full without also adding a complete new set of item cards. While the converse is true of feature cards, there are likely to be fewer features than items. Classification of features is more important than classification of items, and an index can take care of the latter. If items are added seriatim, clusters of holes of particular parts of a feature card may be evidence of correlation. Random distribution of features is also clear to the eye, and it will yield evidence of correlation. If a significant number of items show through holes when two cards are superimposed, there is evidence of correlation between the two features.

Jolley has tried to bring out the versatility, economy, and simplicity of his use of the Carter-Parratt cards, and he has made a convincing case on the basis of the experience of his firm.

Holmstrom, who is working with Jolley and Partners on technical improvement of feature cards, states in correspondence with the writer that he has invented an apparatus to overcome the two main disadvantages of feature cards, viz., 1. the time required for putting feature cards back in a file after removing them for search, and 2. the fact that the searching operation must be repeated on a separate set of cards each time the capacity of one set is exceeded. Details of his invention are not available, although he has deposited provisional specifications both with Jolley and with the British Patent Office.

11. The United States

11. In the United States two federal government agencies, the Atomic Energy Commission and the Office of Basic Instrumentation of the National Bureau of Standards, have made effective and imaginative use of feature cards. A commercial firm, Documentation Incorporated, Washington, D. C., has made important contributions to the mechanization of feature cards with

the aid of federal subsidies.

The Technical Information Service of the Atomic Energy Commission has developed a set of feature cards to ascertain easily and quickly which nuclides possess specified combinations of properties (57, 58). IBM cards were used on account of the relatively large number of punching positions (800) on each card and because they can be punched and collated for distribution with automatic machinery. The complete index includes two sets of cards, each in a different color and each indexing about half of the total population of nuclides (some 1, 200 in all). Each nuclide is assigned a punching position, and there are cards for each property. When the cards are superimposed, the through holes identify nuclides which have the properties represented by the cards at hand. Since there are only 200-300 properties, the number of cards needed is less than half the number needed for an edge-notched system proposed for the same purpose in ORNL 883 (1951), since an edge-notched system would have required a card for each nuclide. Three columns not assigned to nuclides are conveniently used for collating the cards for distribution. A special advantage of feature cards in this situation is that it is unnecessary to refer to an index to identify a punched position with a nuclide: Element symbols are printed above each row (but below row one, on either side of the spaces assigned to the element) and the number of the column, which is its neutron number (indicated by numbers across the top and bottom of the card). There is a special device to aid in the refiling procedure. The group of cards behind each guide card is notched at a different point along the top edge of each card, and misfiling is readily noted by any break in the groove formed by the notches. It is claimed that this set of cards is cheap to use and provides for unlimited expansion. When new data are available, pertinent new property cards can be issued and old ones destroyed. New property cards can be added at will. Any kind of information about nuclides can be indexed. The sets are produced in quantity and distributed to interested research points.

The National Bureau of Standards' Office of Basic

Instrumentation, of which the chief is William A. Wildhack, was set up with the purpose, inter alia, of providing U.S. government laboratories and scientists working on government contracts "with more complete access to existing information on measuring instruments, controls, and data-handling devices" (31). Wildhack, working with Joshua Stern of this office, has made substantial progress in the use and development of possibilities of feature cards. Their objective has been to minimize problems of indexing, storing, and searching information. A basic problem lies in the fact that instrumentation literature is actually a part of many other fields and is not organized anywhere as a separate field (60). Hence the Office of Basic Instrumentation was compelled to develop its own system for collection, organization, and retrieval of pertinent references. One important consideration in indexing the literature of a heterogenous field in which no one is an authority is to make sure that a common language is used by the document analyst and subsequent searchers, and the Office of Basic Instrumentation believes it achieved this objective by setting up ten basic categories, representing major points of view of potential searchers (60, 61). Under each category there is a varying number of primary terms, from ten to 300, and reference terms (synonyms). The feature cards are filed by category and alphabetically within each category.

The Office of Basic Instrumentation found that it had some 10,000-15,000 references to index every year. For this purpose it developed 5" x 8" vinylite cards with 180 columns in 100 rows to provide for 18,000 references. Thus one series of cards will last for at least a year under present publication conditions; and when its punchable positions are exhausted, a new series must be initiated. The holes are necessarily quite small: 0.025" in diameter, and spaced approximately 0.040" on centers. Since no American firm manufactures feature cards as such with perforators and readout devices, the Office of Basic Instrumentation had to develop its own equipment (61).

The level of retrieval of information is a problem of a large proportion of indexing systems, since rela-

tively few actually give the information wanted. The Office of Basic Instrumentation planned to provide abstracts and consider the retrieval process complete when they were delivered (61). If the abstract were available simultaneously with identification of the reference, then much time could be saved by the searchers. Looking toward this end, the Office of Basic Instrumentation has developed microcite on an experimental basis (47, 61). Abstracts are photographed in the appropriate reduction on a matrix film, exactly the size of the feature cards, and the photograph is located in precisely the same area covered by the hole punched for the document which it describes. The full area of the film may be used, since the exposures, unlike the holes, require no supporting area between them. The abstract is read through a microscope or magnifier. The experimental model is based on a card with 1,000 positions, and 3" x 5" typed slips are photographed in a 30 to 1 reduction. If no higher filming ratios are used, extension of the microcite principle to the main sets will require twelve film matrices for each set. However, another suggested development (61) might permit the use of the holes simply to locate the corresponding microphotographed area, not to illuminate the area. Microcite is essentially a return to the primitive efforts of Taylor (56) and the mineralogists (16, 17, 18, 21, 28) to eliminate the rôle of the punched position as an intermediary to get at the needed information and to provide it directly.

Wildhack, Stern, and Smith (61) point out the versatility of feature cards, a quality which Jolley (33) and others have emphasized. At the Office of Basic Instrumentation it is believed that it may be useful for searchers to know which references correspond to the coincidence of a given number of terms that describe the question and which references include a further term not so definitely implied by the question. Again, manipulation of the cards may allow statistical analysis of the literature. A mere count of the holes on a given card (e.g., electromagnetic flowmeters) will give some notion of the extent to which the subject is handled. The number of these that deal with the theory of operation of these devices may be ascertained merely by counting

the through holes when the theory card is put over the first card.

While the Office of Basic Instrumentation's rather simple mechanical procedures will not be reviewed in detail here, it is of interest to note that they report no trouble in adequate alignment of their very small holes. The cards may expand and contract with humidity changes, but "the differential change with humidity from card to card will be much less than the change itself" (61).

Working under contracts with the Armed Forces Technical Information Agency (ASTIA) and the National Science Foundation, Taube and his colleagues (55, V. 2 and 4) have attempted to mechanize and expand the feature card principle in order to handle large collections of documents mechanically. Their work is based on theoretical considerations (V. 2, "The Mechanization of Coordinate Indexing," and V. 4, "Superimposed Coding for Data Storage with an Appendix of Dropping Fraction Tables"); and a rather crude model of a selection and readout machine has also been developed (55, V. 4, "The Prototype Mechanical Alpha-Matrex Machine"). The ideal of an information retrieval system is simultaneous scanning and instantaneous retrieval. Edge-notched and feature cards provide this advantage in a limited way, and the Taube group was interested in expanding the scope and efficiency (defined as a combination of cost and effectiveness of retrieval of information) of feature cards.

The Taube group, like nearly all documentalists who have worked with feature cards, was concerned with the fact that all positions on a given card were dedicated to a specific group of documents and that the number of positions was limited. Moreover, the possible size of the pack of feature cards and its permanence (i.e., the necessity of adding new sheets for new positions) were seen as obstacles to mechanization. In order to overcome these problems, the Taube group proposed (55, V. 2) the creation of artificial alphabets. Feature cards were to be dedicated to letters of these artificial alphabets instead of to vocabulary terms. Thus, if we have 260 letters or ten alphabets, we can express

any word up to ten letters in any Roman alphabet language. Accordingly, all items on accuracy, airplane, acid, accelerometers, etc., will be posted on the A1 sheet; all on acid, accuracy, accelerometers, etc., on the C2 sheet; all on accuracy and accelerometers on the C3 sheet; all on airplane and accelerometer L5 sheet, and so on. Thus the number of feature cards is reduced, space on them is used more economically (about a third of all positions), and the size of the file is stabilized.

While this as yet untested device will provide for more economical use of feature cards, it increases the possibility that two or more entries will overlap. Accordingly, we have the element of "superimposition noise," in which unwanted intelligence may be delivered along with the wanted. Thus, in designing a mechanical selection device based on superimposition, the Taube group had to determine the dropping fraction (i. e., that portion of the whole memory field yielded by a random search of the field) most appropriate to a particular type of coding field and size of collection. To ascertain the average number of false drops per search, we multiply the number of items in a collection by the dropping fraction. To this end the Taube group has presented special and general formulas for dropping fractions and tables for the different sizes of coding fields (55, V. 4). With these tables, they claim, it is possible to select a coding field large enough, and a degree of superimposition so controlled that the dropping fraction is acceptable.

To implement these ideas the group developed a rather primitive type of machine, the Alpha-Matrex (55, V. 4). Cards measuring 16" x 17" were devised to carry 10,000 positions generously spaced at ten 1/16" holes per inch, although a model with a capacity of 40,000 positions is also described. Like Cordonnier (8, 50) and Wildhack and associates (61), the selection of material for the cards (in this case, a specially laminated plastic material) was a problem requiring much study. Other equipment, including the input drill fixture, the selection mechanism, and the readout fixture, were constructed as part of the project. The selection

is implemented simply by typing out the search words on a seven-row keyboard of twenty-two characters (with JK, PQ, and XYZ combined) each; and after the cards are selected, they are placed in a readout fixture behind a translucent plastic raster with 10,000 squares.

Since the Alpha-Matrex machine accepts any notation that can be made with letters (or, in contemplated later models, with numbers), a much larger number of vocabulary terms can be stored here than in purely manual feature cards, Uniterm cards, or conventional catalogs, and there is no need for "satellite catalogs" for authors, sources, projects, etc. All terms are, of course, in seven letters, either expanded thereto by repetition (e.g., air = airaira) or contracted (e.g., acceler = accelerometer). No cases of unwanted answers were reported in service testing.

The Taube group claims that the Alpha-Matrex machine is faster in presentation of answers than any other device, including the highest-speed electronic sequential searching devices; that, for its range of capacity, bit storage costs are well below those of any other machine for information retrieval; that it has unusual freedom from mechanical failure. It is recognized that Alpha-Matrex is only an intermediate device for information retrieval, as contrasted with microcite, Eastman's experimental minicards, and an IBM card carrying an abstract. The developers of Alpha-Matrex say that the most unsatisfactory aspect of the machine is that more intelligence must be applied to the search problems to achieve maximum effectiveness.

12. The Germanies

12. Although the Germans had no part in the early development of feature cards, they have used them more extensively than any other national group; and the recorded applications of feature cards in the Germanies reveal a wide variety of uses.

Most of the reported instances of the use of feature cards are in West Germany. In East Germany fea-

ture cards with 3,500 and 7,000 positions are produced by the VEB Organisationsmittel-Verlag in Leipzig and are used in hospitals and factories (communication from the Deutsche Bücherei). The only library known to use feature cards is that of the Hochschule für Elektrotechnik in Ilmenau (Thuringia), but details of this application are not available. A communication from the Vsesoiuznaia Gosudarstvennaia Biblioteka Inostrannoi Literatury in Moscow indicates that feature cards are known and have been studied in the Soviet Union and the people's democracies, but so far no library is known to have used them.

Two West German office supply firms produce widely used feature cards. Allform, Brandenburgische Strasse 27, Berlin W. 5, advertises three types of cards and a simple punching apparatus (2), and there is a full account of their practical application in Jaeckle's Wirtschafts-Praxis (32). For comparatively small collections of documents Allform offers a card with 2,000 positions, divided into twenty squares of 100 positions. For larger collections there is a 6,000 position card. In addition, there is a card for chronological indexing over a period of five years. Thirty-one positions (the maximum number of days in a month) are provided in each vertical column. The twelve vertical columns total 372 positions for one year, 1,860 for five. Thus not only correspondence but also daily collections of data can be indexed by subject. At the top of all Allform cards there are twenty-five positions for the letters of the alphabet (xy being one unit) to be notched to show at a glance the alphabetical position of the particular card; and on the last fifth of the upper edge there are six numbered positions to be notched to show the series. Thus any misfiled card can readily be noted. Allform cards are marketed in the Netherlands under the name of Transelecta (45).

Edler und Krische, Kestnerstrasse 42, Hannover, offer Ekaha feature cards and a simple punching apparatus (2). The Ekaha cards have 7,000 positions and are twice as large as the 6,000 position Allform cards, with correspondingly larger squares for punching. The cards come in several colors for different series.

Braband (7) describes the special applications (although not referring to Ekaha cards by name) and places particular emphasis on their use in searching paten literature.

Edler und Krische have printed a somewhat different type of card that is used by the Deutsches Kunststoff-Institut, Schlossgartenstrasse 6 R, Darmstadt (39, 40). The cards have 6,000 positions and are divided into sixty squares of 100 positions with numerals from 0 to 59 lightly printed over them. The intersections of the lines, not the squares, as in standard Ekaha cards, are punched. Like the Allform cards, this type has an alphabet at the top to be notched as a filing aid. Similarly, a group of numbers from 1 to 10 on either side of the card could serve as identification of the series to which the card belongs. The Kunststoff-Institut maintains an abstract file with a number corresponding to the document's numerical position on the feature card, and there is also an alphabetical author index. Knappe states that the classification by which the feature cards are arranged is being constantly refined (39), and his problem is in general rather similar to Batten's (4) in this respect. As a special virtue of feature cards as used in the Kunststoff-Institut, Knappe (40) says that the time required for searching does not increase in proportion to the increase in the number of documents, while this factor remains constant in the case of edge-notched and machine sorted cards.

German scientists have put feature cards to effective use in a wide variety of situations. Martin Scheele, a limnologist, devoted a substantial part of his study of punched cards to the Sélecto system, using his own field as an example (48). He cites as a special quality of feature cards the fact that additional documents can be added at any time and subjected to the full range of subject analysis already at hand, or to new subjects, if necessary. He recommended Hollerith (IBM) or Powers cards (with 960 and 540 positions respectively) for limited collections, and he points out that the Powers cards are especially handy due to the availability of an inexpensive hand punch. Karl Eduard Rotschuh, a Munich physiologist, has followed the same line and emphasized

the utility of feature cards for reprint collections (47a). He developed his own cards for the 3,700 reprints already in his collection and the some 3,000 more he expects to acquire. He divided his cards into squares of 100 positions assigned to various letters of the alphabet according to their frequency as initials of the last names of authors already in his collection. He assigned 350 catchwords to feature cards and arranged them according to a simple home-made classification. Kistermann and Uhlein (37, 38) recommend feature cards for essentially the same type of collection of references. They recommend edge-notching of the feature cards as a convenient selection aid by which the desired cards may be pulled from the pack (37; also recommended by Holmstrom, 22).

The Germans have used feature cards widely in various aspects of medical research and medical records. Scheele placed heavy emphasis on this application. Adler (1) applied Allform cards to pharmaceutical research. Udo Derbolowsky, a Hamburg physician, has written extensively on feature cards in medical research (11, 12, 13, 14). In an earlier article (15) he had recommended machine searching for patent records, recognizing clearly the relatively high cost of this method; but later he showed how essentially the same information could be retrieved by use of Allform feature cards (13, 14). At the same time he demonstrates the documentary applications of feature cards (13, 14). He recognizes the superior utility of Hollerith (IBM) cards for tabulation and correlation, but he advocates the use of Sélecto cards for limited literature searches (12). In a chiropractic study of pelvic movements he recommends Sélecto cards as a medium for recording observations, but he goes into no details on the specific application (11).

While the Germans have demonstrated the large variety of uses of feature cards, only Heinze (25a) has done any imaginative, original work to refine and expand the applications of feature cards, and even he has limited his work to indexing journal articles. On a mathematical basis Heinze demonstrates the need for cutting down on clerical work in compiling references

and at the same time recording all subjects of potential interest on the appropriate cards. He is particularly eager to eliminate the copying of exact bibliographical references in full and the waste of space once the maximum number of references has been attained on a feature card. He overcomes both of these handicaps by punching positions that correspond to page numbers of articles instead of giving each reference a serial number and punching the corresponding hole. He allots 1,000 positions each year to each feature card for this purpose, and he writes the abbreviation of the periodical in question beside each perforation (preceded by 1, 2, etc. if the page number is greater than 1,000, 2,000, etc.). If, by chance, two or more articles on the same subject appeared in different journals in the same year with the same page number, an auxiliary card of a different color could be used. Incidental advantages of this device are to show the volume of relevant literature published in any one year and the possibility of studying obsolescence of references by checking or ringing the holes with a different color of ink representing the year in which they were consulted.

Heinze promises a subsequent article in which he will examine the possibilities for photo-electric selection of feature cards. Here is a possible approach to the solution of the search problem which the Taube group views as the most serious barrier to maximum effectiveness in the use of feature cards (55, V. 4).

2. Horace Taylor's Patent

2. Taylor's "selective device" (56) is claimed to identify "a particular species or a particular number of a particular species... with ease and accuracy." His illustrations in the original patent show an adequate sample of New England birds. There is no evidence that this system is generally effective for ornithological taxonomy. This fallacy is implicit in many of the essays on specific applications of feature cards and will not be recited in detail, although the critical reader should bear in mind that much of the literature on feature cards is aimed at individual applications. Only the Taube group has given appropriate attention to the dropping fraction (55, V. 4) and the problem of unwanted intelligence that may be yielded along with the wanted.

Taylor says his cards are "so arranged" that the screen sheets may be turned to four positions (recto, top and bottom; verso, top and bottom) to be applied to four different base sheets, but his cautious patentese style refrains from disclosing the secret of this important trick. The method of his quadriform arrangement should be disclosed; and, if generally applicable to feature card systems, it should be refined.

3. H. E. Soper's Patent

3. Soper (52) gives examples of classification of individuals, for instance, by occupational class or health grade, as examples of the compilation of "tabular and statistical data" on his feature cards. His broad claim of superiority "over systems of indexing employing cross references or multiple entry" is not supported. Subsequent publications provide evidence in favor of this idea in specific situations, but no one has

established the general validity of Soper's claim. This matter should be a key point in consideration of specific new applications of feature cards.

5. Feature Cards for Mineral Identification

5. Gray (25), Donnay (16, 17), and Fairbanks (21) do not argue for the infallibility of their feature cards, but they say that they are superior to conventional tables for mineral identification. In each case they cite examples of the inadequacies of tables, although all but one article (17) is conservative in claims for the infallibility of feature cards in mineral identification.

A much more fundamental problem is involved in Hurlbut's claim for the superiority of edge-notched cards in mineral identification over the feature cards developed by Donnay. Hurlbut's argument that one must thumb through a whole pack of feature cards to find the properties for which one is searching is not supported. Batten (4), Cordonnier (8), the Institut des Fruits et Agrumes Coloniaux (29, 30), and others have proven that adequate classification obviates the need for unsystematic search for needed feature cards in a large group. Holmstrom (22) and Kistermann and Uhlein (37) suggest edge-notching of feature cards for preliminary selection for searching cards in a large group. Donnay (16) separated his relatively small group of feature cards with buff tabs. The Synoptic system of tabs developed and marketed by B. Lampel (41, 51) is another aid to selection of feature cards for searching and was advocated by Loosjes in conjunction with Uniterm (43, 44, 45). Systematic notching to form a regular pattern when the cards are filed correctly was advocated by Wachtel (58). Hurlbut's statement that his cards can be thrown back together in any arrangement after use is significant. This is a basic advantage of edge-notched cards. If the yet undeveloped suggestions of photo-electric selection from feature cards (4, 25a, 27), find a practical solution, the general application of Hurlbut's statement will be invalidated. In any event, we need comparative time studies of the use of feature cards and

of other systems comparable in simplicity, scope, and function.

6. W. E. Batten's Feature Cards

6. Batten's statement that no mechanical aid will make up for a defective classification system is borne out in his particular case by the detailed account of his classification of the patent literature of plastics (4) and the note in the F.I.D. Manual (22, 744.432 E 1). However, this statement needs to be modified by further investigation of Loosjes' proposals for the use of Uniterm in connection with feature cards (43, 44, 45), by the theoretical work of the Taube group (55, V. 2 and 4), and possibly also by as yet unpublished ideas of Holmstrom (correspondence with writer) and Heinze (25a).

Holmstrom (22) does not develop his suggestions for the "facets" in Ranganathan's Colon Classification in connection with feature cards, but it seems to be worth further study.

Batten's doubts about the wider application of his system may have stemmed from concentration on his own problems. Literature which will be discussed later (e.g., 8, 25a, 45, 54, 55) will indicate far wider applications than Batten envisioned.

Batten's statement that technical man-hours are saved but not clerical man-hours is obvious. However, we need to examine the cost elements in each documentation project to ascertain the economical limits of the expenditure of clerical man-hours in the use of feature cards.

If the number of searchable features is unlimited (27), we need evidence to point out the economic limitations of the number of feature cards in a single series (cf. 55, V. 2). It is possible that there is no limit but it would be hazardous to plan such a file without some consideration of the facility with which it can be handled.

Batten's passing reference to feature strips may deserve further study. Would it prove to be an improvement or an onerous complication to a mechanized selection system?

7. G. Cordonnier and Developments in France

7. The 12,500-position Cordonnier card is said to be large enough to allow the introduction of new documents by leaving a certain number of positions empty (22). This statement is made with reference to collections not likely to exceed 12,500 items. The constant accretion of documents is as serious a problem for feature card systems in general as the expansion of subject content is for edge-notched cards.

There is no statement in the literature about the actual steps needed to protect feature cards from expansion and contraction. The formula for Sélecto plastic sheets and cellulose cards is not given. Other possible dangers, e.g., mold and insects in the tropical stations of agencies such as the Institut des Fruits et Agrumes Coloniaux, are not mentioned. This writer has seen a shipment of 100,000 all-rag catalog cards reduced to dust by the polillas of Martinique after a week in the custom house.

The effectiveness of the various applications of Cordonnier's cards is mentioned (22), but no comparative studies contrasting feature cards with other systems of selection, have actually been conducted. The Taube group outlines the elements of such a study (55, V. 4).

There must be some awkward peculiarity about the French judicial identification system to require the use of feature cards for fingerprint identification. A communication from the Identification Division of the Federal Bureau of Investigation in the writer's file advises that this agency has tried punched cards (but not feature cards) for fingerprint searching and found this method impractical. If feature cards are practical for fingerprint searching, this application needs to be studied in detail.

The Institut des Fruits et Agrumes Coloniaux (30) believes that its colonial branches can depend on the feature cards for selecting references they need. No evidence other than the continued use of the system is available. Holmstrom (22) says it saves the expenses and delay of compiling indexes to several different bibliographical organs. It would be well to know the names of these organs as a case study in what might be clumsy bibliographical organization.

In connection with Holmstrom's list of apparent objections (22), it would be well to document some of his refutations with comparative cost studies of different selection systems under the same conditions.

The suggestion of edge-notching feature cards for selection from the file (22, 37) needs to be developed and described in detail.

Holmstrom (22) says a complete search of any general subject can be made. He proves this statement by showing how an effective classification, indexed properly by the arbitrary serial numbers of the cards, can theoretically provide all possible subjects to be searched.

8. The Netherlands

8. Westendorp (59) brings out the contrasts between optical and machine searching (or simultaneous and sequential searching) of the same type of card and says that machine searching permits the search for several subjects simultaneously. It is apparent that this need rises from his specific problems (probably of computation and tabulation) at Philips, although he does not say so specifically. A broader comparison between the two systems in situations where both might be used would be useful. Westendorp also states that machine sorted cards wear out more rapidly than do feature cards. We need a study of the life expectancy of cards under both systems, and suitably durable material should be developed to avoid replacement. It is possible that cards developed by Cordonnier (8, 50), the Office of Basic Instrumentation (61), and Taube and his col-

leagues (55, V. 4, "The Prototype Mechanical Alpha-Matrex Machine") should be tested for durability under manual and mechanical conditions.

Loosjes (43, 44, 45) makes strong claims for the superior "readability" of Delta cards. Since he has also used Sphinxo cards, these claims are presumably based on his personal experience. Loosjes' claims suggest the need of a study of the rapidity and ease with which different varieties of feature cards may be read.

In only one reference (45) does Loosjes expand on his advocacy (43, 44, 45) of the use of Uniterm and the Synoptic filing system in connection with the use of feature cards. He provides no more evidence for these contentions than what is quoted in this text; and his recommendations for the use of Uniterm must be supplemented by reference to the work of Taube and his associates (55, V. 2, "The Mechanization of Coordinate Indexing," and V. 4, "The Prototype Mechanical Alpha-Matrex Machine").

If feature cards record the dynamics of scientific research more effectively than other systems (43, 45), we need to know specific types of research of which the dynamics need to be expressed. The enthusiasts for feature cards see the virtues of this system as applied to their own problems, but the comparable utility of feature cards, edge-notched cards, linked hole cards, and machine selected cards must be closely defined.

10. England

10. The claims made in the Carter-Parratt promotional folder (7a) are supported only in part by evidence. Holmstrom's communication with the writer dealing with his invention may make the claim for high speed of operation true in all cases, but it is not true at present except for systems involving only one or two series of feature cards. Elsewhere (22) Holmstrom says several different desks are needed to operate some systems with multiple series. It is open to question

whether the speed thus obtained by human labor is more economical than expensive mechanical equipment when we depart from Batten's contention (3, 4) that feature cards are for the small operator. There is a tempting (and possibly misleading) analogy here between the arguments between devotees of the abacus and the high-speed digital computer.

If there is a constant control of data, it is at the cost of time needed to refile cards (cf. Holmstrom's invention, of which we have no details as yet).

The simplicity and economy of feature cards have been brought out by all users. Economy of large-scale systems involving two or more series of cards, is subject to investigation.

Compactness is a virtue until the point at which features may proliferate rapidly. Conceivably a few thousand items might have to be indexed under a disproportionately large number of features.

Feature cards are versatile for the applications recommended by Jolley and Carter-Parratt (7a, 33, 34, 35, 36). Application to a complex body of scholarly literature might cancel this claim in some instances. It might be well to attempt to apply feature cards to some bodies of literature in which their utility will be much less than in fields where they have been used thus far. A few negative results would help to define the limits of their practical utility.

Jolley (33) offers an extreme case of the comparison of feature cards and item cards. In his application to personnel records (34), he is dealing with a known number of employees (items) in a given agency, and features needed to describe personnel for management purposes are not excessively numerous (34). Universal application of this argument to all possible uses of feature cards is subject to investigation in each instance. Jolley's suggestion of the use of feature cards for correlations (34) is significant. He says (34): "In research applications this means that no line of inquiry need be followed up unless there is already evi-

dence that a correlation of one sort or the other exists between the relevant features."

11. The United States

11. Wachtel's argument (57, 58) for her feature cards over edge-notched cards is supported in part by the simple arithmetic that the latter require twice as many cards in a pack as the former. She does not prove convincingly that her edge-notching device for refiling is a substitute for the advantage of edge-notched cards that they can be thrown back together at random (cf. Hurlbut, 28).

"Unlimited expansion" of the Atomic Energy Commission cards needs definition. It is unlikely that any substantial number of new properties will be identified. Here, as in the case of the mineralogists' cards, we have a relatively stable number of features and items; and therefore claims made for these two systems are not always universally applicable.

The level of retrieval has been set by the operators of the Office of Basic Instrumentation system as the provision of abstracts. It is possible that this is satisfactory to the laboratory worker in this field, but evidence is lacking. If microcite and its potentials are fully developed, the desirable level of retrieval for each case needs to be identified.

Substantially larger surfaces than 3" x 5" cards have been legibly reproduced on film on areas even less than 0.025" in diameter. This possibility needs further study with relation to its applicability to the microcite principle. Possible relevance of the minicard system also seems to justify study.

Wildhack et al. (61) argue for the versatility of feature cards. As examples, they cite the possibility of identifying some less important or less obvious feature of a question and the statistical analysis of the literature. Together with evidence cited by Jolley (33), there is good reason to accept the idea of the versatil-

ity of feature cards. However, the various aspects of this versatility should be identified, and the applicability of feature cards to individual classes of situations should be made clear.

The practical value of the dropping fraction tables (55, V. 4) is illustrated by four examples. No large scale operations are analyzed, but there is no reason to assume that the hypothetical examples are not typical.

Just as in the case of the special cards made up by Cordonnier (8, 50) and the Office of Basic Instrumentation (61), it would be desirable to know the precise composition of the cards. Taube's group says that "fabrication of the cards turned out to be the most critical aspect of the entire project."

The claim for the speed of search by the Alpha-Matrex machine is supported by actual time studies, with tabulations. Comparative studies are not made, but an outline for such a comparative study is presented. The relative expense of the machine is as self-evident as is the relative cost of manual feature card sets and machine selection devices. As for the freedom from mechanical failure, the inventors suggest that Alpha-Matrex may be deceptive of the facts of machine life and become more typical as it becomes more complex. As for the intermediary character of the Alpha-Matrex, the suggestive work on microcite and minicards suggests the need and feasibility of serious investigations along these lines.

Jonker Business Machines, Inc., 404 North Frederick Avenue, Gaithersburg, Maryland, headed by Frederick Jonker, a former Taube associate, has been working on practical developments of Alpha-Matrex. Mr. Jonker has developed and marketed a "Termatrex" machine to handle cards with a basic capacity of 10,000 items of data. He is now working on a system that "will search millions in a matter of minutes," completely automatic push-button equipment, and equipment tie-in with computers and IBM systems. He is fully aware of the need for direct access to information rather than using the Termatrex machine as an intermediate device,

and his firm is working on solutions for these problems.

12. The Germanies

12. Knappe (40) demonstrates on graphs the results of actual time studies made to prove his contention that time required for searching a feature card system does not increase in proportion to the increase in the number of documents.

Derbolowsky (12) says "Die Hollerith-Methode ist auf medizinischem Gebiet die Methode der Korrelations- und Grosszahlenforschung," without further explanation. Jolley (33) has shown that certain types of correlations may be readily identified by feature cards, e. g., clusters of perforations at one point on a card when documents are recorded chronologically, or a significantly large number of through holes when two cards are superimposed. This is a useful example of the type of information that needs to be developed and critically examined in a comparative study of various systems of punched cards which are put to the same tasks.

Heinze (25a) gives no indication of the character of his proposed photoelectric device for selecting cards for examination. Here is a potentially significant contribution. The comparative efficiency of this device, of Holmstrom's yet unrevealed notion for simplifying the searching of multiple sets, and of the Alpha-Matrex will need to be subjected to detailed comparison in terms of expense and effective retrieval of information.

Major Targets for Research in Feature Cards

The disarming simplicity of the basic feature card principle tempts us to overlook broad elements of their efficiency (cost of equipment and operation and effectiveness of information retrieval). Only the Taube group (55, V. 4) and Heinze (27a) have given proper attention to this element. Detailed time and cost studies, starting with the outlines of the Taube group, must be made in a large number of characteristic situations before we can accept all the claims made for their efficiency (as defined above). In particular we need comparative studies with other sequential and instantaneous information retrieval devices, both manual and mechanical, and with the various types of library catalogs (author, subject, dictionary, classed, card, and printed).

Some of the enthusiasts for feature cards have been a bit reckless in their claims for the versatility and wide application of the system. The limits of utility of feature cards, both for subjects and for methods of investigation, need to be closely defined. This may best be achieved in connection with studies proposed in the preceding paragraph.

The mechanization of feature cards, actually achieved by the Taube group and seriously proposed by Batten, Holmstrom, and Heinze, is the next major step forward. At the same time the search problem must be approached in the same imaginative terms as those applied by the Taube group and Heinze. Symbolic logic as used by the Taube group will carry us only so far. Rigid tests of feasibility are indispensable at each stage of theoretical or mechanical development.

The level of retrieval of information is a grave problem in all systems. Rider attempted to make it absolute by inserting microcards with full texts in library

catalogs in lieu of author cards. Where little information is needed (e.g., the cards on properties of nuclides), such a solution is easy. We need to know how acessible full or partial information must be to workers in various fields, how economical it is to provide it, and in what form it should be given in different fields. To what extent is microcite the answer in a feature card system? Can minicards be adapted to a feature card system? What photographic reductions are feasible in such cases?

How practical is it to "publish" feature card packs such as those on properties of nuclides or those of the Institut des Fruits et Agrumes Coloniaux? To what situations can such "publication" be applied?

Apparently the Taube group, Cordonnier, and the Office of Basic Instrumentation have overcome the problems of producing cards to hold pinpoint perforations accurately and, in the first case, to stand up under machine handling. This problem will remain with us if we move toward cards with extremely large numbers of positions and more machine handling. It is a basic practical problem that will require constant attention.

References

1. Adler, Fritz Helmut. Die sichtlochkartei -- ein vieldimensionales inhaltsregister. Die pharmazeutische industrie, XIX (1957), 170-2.
2. Allform-Nummernkarte. Berlin, Allform Büroorganisation, n. d. (Promotional leaflet).
3. Batten, W. E. A punched card system of indexing to meet special requirements. Association of special libraries and information bureaux, Proceedings of the twenty-second conference, 1947, 37-8.
4. Batten, W. E. Specialized files for patent searching. In Robert S. Casey and James W. Perry, eds., Punched cards, their applications to science and industry (New York, Reinhold, 1951), 69-81. (1958 edition unavailable at time of writing.
5. Björkbom, Carl. Titthålkort. Teknisk dokumentation, XIII (1957), 54.
6. Borgeaud et cie. and Henri Liber. Système et appareil de classement par feuilles mobiles. French Patent 565,475. 28 January 1924.
7. Braband, Carl. Einige vorschläge für die ausgestaltung des sichtlochkartenverfahrens. Nachrichten für dokumentation, VIII (1957), 42-44.

7a. Brisch-Vistem data processing system. London, Carter-Parratt Ltd., n. d. (Promotional folder).

8. Cordonnier, G. Classification, classement, rangement et selèction. CNOF [Comité National de l'Organisation Française], XXV (April, 1951), 3-8, and XXV (May-June, 1951), 19-35.
9. Correspondence regarding metallurgical documentation and the Cordonnier-Batten system of punched cards. Monthly Circular for members of the UNESCO committees and collaborating bodies concerned with the documentation of the natural

sciences, No. 6, 15 November 1952, 6-11.
10. Delta-Kaarten. The Hague, Semper Avanti, n.d. (Promotional brochure).
11. Derbolowsky, Udo. Beckenmechanik -- chiropraktisch geschen. Hippokrates, XXVII (1956), 310-3.
12. Derbolowsky, Udo. Das geistige quantitativproblem in der medizin und die mechanisierte dokumentation als lösungsversuch. Klinische wochenschrift (1953), 825-31.
13. Derbolowsky, Udo. Die sichtlochkartei als nothelfer des arztes. Münchener medizinische wochenschrift, XCIX (1957), 1547-52.
14. Derbolowsky, Udo. Die sichtlochkartenverfahren in der medizin. Hippokrates, XXVIII (1957), 469-73.
15. Derbolowsky, Udo. Zur mechanisierung der dokumentation auf dem gebiet der medizinischen psychologie. Nachrichten für dokumentation, III (1952), 23-24.
16. Donnay, J.D.H. A small set of grids for the determination of non-opaque minerals. American mineralogist, XXIII (1938), 91-100.
17. Donnay, J.D.H. Systeme de grilles pour la determination des mineraux. Annales de la Societe geologique de Belgique, LIX (1935/6), 250-7. English abstract in mineralogical abstracts, VI (1957), 431.
18. Donnay, J.D.H., and J. Melon. A system of grids for the determination of non-opaque minerals. American mineralogist, XXII (1937), 218.
19. Draheim, Heinz. Benennungen von lochkarten; ein beltrag zur entwicklung einer einheitlichen terminologie der lochkartenverfahren. Nachrichten fur dokumentation, VIII (1957), 94-7.
20. Ekaha-Sicht-Lochkartei. Hannover, Ekaha-organisation, Edler & Krische, n.d. (Promotional leaflet).
21. Fairbanks, Ernest E. The punched card identification of ore minerals. Economic geology, XLI (1946), 761-8.
22. F.I.D. Manual on document reproduction and selection. The Hague, F.I.D., n.d. (Looseleaf; see sections 744.43 E1 et seq.; unsigned; authorship attributed to J. Edwin Holmstrom in 54).

23. Findex, en ny form av referenskartotek. Stockholm, Esselte, n.d. (Promotional brochure).
24. Findex -- en ny form av referenskartotek. Effektivitet, Tidskrift fur presentation av Arbetsforenklande Kontorstekniska Hjalpmedel, No. 2, 1957, 2.
25. Gray, C. J. A new method for using physical characteristics of minerals for their identification. Transactions of the Geological society of South Africa, XXIII, 1920 (1921), 114-7.
25a. Heinze, Helmut. Rentabilitatsbetrachtung einer schriftlumskartei. Nachrichten fur dokumentation VIII (1957), 125-8. English abstract in UNESCO Monthly bulletin on scientific documentation and terminology, 1958.03, 15 March 1958, paragraph 7.
26. Holmstrom J. Edwin. Documentation in the fisheries biology branch of the Food and agriculture organization. UNESCO Monthly bulletin on scientific documentation and terminology, 1957.02, 15 February 1957, paragraph 3.
27. Holmstrom, J. Edwin. Facts, files and action in business and public affairs. Part two: Filing, indexing and circulation. London, Chapman and Hall, 1953.
28. Hurlbut, C. S., Jr. Mineral determination by means of punched cards. American mineralogist XXXIII (1948), 508-12.
29. Institut des fruits et agrumes coloniaux, Paris. La code de classification, Paris, n.d. 19 p. Mimeographed.
30. Institut des fruits et agrumes coloniaux, Paris. Rapport annuel, 1948. Paris, 1948 (?).
31. Instruments reference service. U.S. National bureau of standards, Technical news bulletin, XXXIX (No. 8, August, 1955), 110-1.
32. Jaeckle, Julius F. Wirtschafts-Praxis. Stuttgart, Forkel-Verlag, 1957. Teil-Lieferung 108, 15 May 1957, Die sichtlochkarte, 6 unnumbered p.
33. Jolley, J. L. Card indexing radically revised. The Times review of industry, XII (1958), 97-98.
34. Jolley, J. L. Personnel records -- a new approach. Feature card system saves time and space. Personnel management and methods, XXIV (No. 209,

November 1957), 20-21.
35. [Jolley, J. L.] Some notes on the use of feature cards for library indexing purposes. London, J. L. Jolley, n.d. 2p. Mimeographed.
36. [Jolley, J. L.] The use of punched feature cards in research. London, J. L. Jolley, n.d. 2p. Mimeographed.
37. Kistermann, Fr., and E. Ühlein. Die sichtlochkarte; vereinfachte auswertung von literatur und untersuchungsergebnissen. Umschau, XII (1957), 370-1.
38. Kistermann, Fr., and E. Ühlein. Die sichtlochkartei; zur technik der wissenschafilichen arbeit. Diskus, 1957, no. 5, Beilage, 185-6.
39. Knappe, Werner. Die dokumentation im Deutschen Kunststoff-Institut. Chemische industrie, X (1956), 511-2.
40. Knappe, Werner. Einige gesichtspunkte zur anwendung von sichtlochkarten. Nachrichten für dokumentation, VII (1956), 140-4. English abstract in UNESCO Monthly bulletin on scientific documentation and terminology, 1956.10, 15 October 1956, paragraph 10.
41. Lampel, B. Une technique nouvelle de classement des fiches et des microfilms. UNESCO Monthly bulletin on scientific documentation and terminology, 1957.09, 15 September 1957, paragraph 21.
42. Lingenberg, Walter. Ueber die Anwendung von lochkartenverfahren in bibliotheken. Cologne, Greven Verlag, 1955. (Arbeiten aus dem bibliothekar-Lehrinstitut des Landes Nordrhein-Westfalen, 9).
43. Loosjes, Th. P. The Delta card. Aslib proceedings., IX (1957), 142-3.
44. Loosjes, Th. P. Die Deltakarte, eine weitere entwicklung der sichtlochkarte. Nachrichten für dokumentation, VIII (1957), 90.
45. Loosjes, Th. P. Documentatie van wetenschappelijke literatuur. Amsterdam, Noord-Hollandsche Uitgeversmij, 1957.
46. Metcalfe, John Wallace. Indexing and subject cataloging. Alphabetical: classified. Coordinate: mechanical. New York, Scarecrow press, 1957.

47. Microcite, an aid to more effective researching. U. S. National bureau of standards, Technical news bulletin, XL (No. 9, September, 1957), 141-2. Also issued in mimeographed form (5 p.) by the Bureau of standards in August, 1957, as Summary technical report 2144.
47a. Rotschuh, Karl Eduard. Eine sichtlochkartenmethode zur erfassung einer grossen sonderdrucksammlung. Nachrichten für dokumentation, IX (1958), 90-1.
48. Scheele, Martin. Die lochkartenverfahren in forschung und dokumentation mit besonderer berücksichtigung der biologie. Stuttgart, E. Schweizerbart'sche verlagsbuchhandlung, (Erwin Nägele), 1954. Reviewed by Karl F. Heumann in American documentation, VII (1956), 330, and by D. J. Campbell in the Journal of documentation, XI (1955), 89.
49. Scheele, Martin. Über einige grundfragen der dokumentation mit lochkarten. Nachrichten fur dokumentation, VI (1955), 111-8.
50. Schürmeyer, Walther. SELECTO -- Ein neues auswahlsystem für die dokumentation. Nachrichten für dokumentation, III (1952), 33-4.
51. Service d'organisation synoptic, Paris. Bulletin technique, No. 1, 1952- .
52. Soper, Herbert Edward. Means for compiling tabular and statistical data. U. S. patent 1, 351, 692. 31 August 1920. Abstract in Official gazette of the U. S. patent office, CCLXXVII (31 August 1920), 928.
53. Sphinxo. Paris, Détectri (68, rue de Richelieu), n. d. (Promotional brochure).
54. "Superimposable," "optical," "peephole," or "coincidentally" punched cards. UNESCO Monthly bulletin on scientific documentation and terminology, 1958. 03, 15 March 1958, paragraphs 6-8. (Probably written by Holmstrom.)
55. Taube, Mortimer, et al. Studies in coordinate indexing. Washington, Documentation incorporated, 1953-1957. 4 v.
56. Taylor, Horace. Selective device. U. S. patent 1, 165, 465. 28 December 1915. Abstract in Official gazette of the U. S. patent office, CCXXI

(28 December 1915), 1078-9.
57. Wachtel, Irma S. Indexing nuclear data on punched cards. Preliminary edition. Oak Ridge, Tenn., Technical information service, U.S. Atomic energy commission, 26 April 1951. (TID-469).
58. Wachtel, Irma S. A punched card index for nuclear data. American documentation, III (19520, 56-7. Reprinted, with slight modifications, in 55, v.2, 62-66.
59. Westendorp, J. De toepassing van ponskaartensystemen bij de documentatie van octrooiliteratuur op het gebied van de chemische technologie bij de octrooiafdeling van de N. V. Philips' gloelampenfabriken en de voor- en nadelen hieraan verbonden. Bibliothekleven, XLI (1956), 289-98.
60. Wildhack, William A., and Joshua Stern. The peek-a-boo system in the field of instrumentation. In Jesse H. Shera, A. Kent, and J. W. Perry, eds., Information systems in documentation (New York, Interscience publishers, 1957; 2v.), II, 209-25.
61. Wildhack, William A., Joshua Stern, and Julian Smith. Documentation in instrumentation. American documentation, V (1954), 223-37. Also issued as National bureau of standards, NBS report 3276 (April, 1954; 25 p.); and reprinted in 55, V.3, 114-156, omitting the list of references and not indicating original place of publication.

Volume Four Part Three

PUNCHED CARDS

by

Ralph Blasingame, Jr.

I. Introduction

The history of the application of machine-sorted punched cards to libraries is a relatively brief one - perhaps twenty-five years at most. During that time, applications have been made in three main areas: first, in the routine, repetitive tasks concerned with circulation, acquisitions, accounting and the like; second, in the somewhat broader and possibly more significant area of bibliographic control; and third, in preparing copy for standard library tools (lists and catalogs). In the first type of application, standard machines have been used. In the field of bibliographic control, on the other hand, a considerable amount of effort has been expended on the development of mechanical punched-card and related equipment particularly applicable to the requirements of the problems encountered. From these two facts relating to equipment development, one may infer that machine-sorted card equipment is not especially profitable in the first type of application (though perhaps the only profit consideration has been on the part of the equipment manufacturers) and that the second type of application has attracted people of greater aggressiveness and with more pressing problems. Both inferences are probably correct, and it is interesting to note that the use of punched-card equipment in the first type has, judging from frequency of mention in the literature, reached a standstill, while the punched-card seems to be headed for an incidental role, perhaps as a programming device, in the second type. The third type appears to have passed the experimental stage, but has not been widely accepted as compared to the first or second.

The customary aim of the first type of application is to free the librarian to work in more important areas, normally in public services. For the second, the general justification is to give a measure of control over the literature. For the third, to make information on

library holdings known over a wide area is the common aim.

Information for this study was gathered through searches in Library Literature, Library Science Abstracts and the "Annotated Bibliography on Uses of Punched Cards" in Punched Cards, Their Applications to Science and Industry (1958), second edition (1). This study is not comprehensive in the sense that every article located has been mentioned. An effort has been made to sort out those items of the literature of the application of machine-sorted cards to libraries which seem to typify applications in the areas discussed. For convenience, IBM and Remington-Rand cards and related equipment and cards will be lumped together hereafter under the term machine-sorted cards.

II. A Routines

As a substitute for manual filing, sorting and record-keeping, machine-sorted card systems date back at least as far as 1936 (2). Soon after that, some attention was given to machine-sorted cards as a device for recording library statistics (3). Growth of the application of these machines to library routines was slow, perhaps because of the depression budgets on which many libraries existed or perhaps because librarians have generally been slow to adopt machine methods. By 1943, however, at least six applications were known (4); by 1946, at least eight were recorded (5); and by 1950, seventeen such applications were located (6). Though by far the greatest activity in this general area has been in American libraries, some work has taken place elsewhere, especially in England (7).

A review of the literature of applications of machine-sorted cards to library routines is an especially unrewarding experience. With a few exceptions, the articles one encounters are speculative, painfully uncritical, repetitive, or some combination of the three. This is not necessarily a criticism of librarians; the seeming magic of punched card machinery has enchanted a good many supposedly hard-headed business men. An

indication of general use of the machines lies in the phenomenal increase in the market price of IBM Corporation stock. Perhaps another indication is the fact that only three articles were found after protracted search which reported that a machine-sorted card application had been discontinued. This latter fact may, of course, also show that converting to a machine-sorted method may be quite the same as taking a bear by the tail.

Interest in applying machine-sorted cards to library routines, judging from the number of articles located for this study, has flagged since the early 1950's. Of the forty-five references examined, only three date from 1955 or later, while twelve appeared between 1950 and 1955 and twelve between 1944 and 1949. Further, those published since 1955 do not describe new applications.

Several reviews of machine-sorted card applications have appeared, none of which furnishes a complete coverage, probably because the literature is scattered through state library association bulletins, master's essays, and pamphlets, in addition to the more or less standard library journals. Perhaps the most complete and up-to-date review has been made by Berry (8). Another of note for its selection of unusual applications rather than completeness is by Gull (9). Parker's descriptive book on punched-card application is required reading for anyone wishing a review of procedures (10).

The following routines are among those to which machine-sorted cards have been applied in some way:

Circulation
Acquisitions and Accounting
Analysis of Book Stock
Serials Acquisition and Control
Preparation of Catalogs
Shelf-listing
Payroll and Personnel Records
Billing
Inventories of Equipment and
  Materials

Circulation

Perhaps because circulation routines present the same relatively simple problems time after time, a comparatively large number of libraries have applied machine-sorted cards to them. The variety of methods used may reveal both the flexibility of use of the machines and the ingenuity of librarians. Methods here may be classified in three general types: first, those which make a mechanical record from book and borrower cards; second, those in which the punched card is used as a call slip and is punched for date due (and sometimes for other information); and third, those in which the card carries only a serial number and which are more or less straight-forward transaction number systems.

Of the first class, the system installed at the Montclair (New Jersey) Public Library is certainly the most publicized (11). This system originally involved a machine which has not been made generally available. However, the principle advantages of it may be achieved by the use of standard machines through a method suggested by Callander (12). The comparative expense, however, has not been determined or, at least, has not been made public. The immediate drawback to the use of this system is that it requires that a card be punched for each cataloged item in the collection and for each borrower. Selected information from the borrower's card and the book card are punched automatically into a third card (remotely, if desired). Upon return of the material, a fourth card is punched and the "out" and "in" cards then constitute the record of each loan.

A description of the second type of circulation system may be found for the University of Florida (13). Related systems are in use in the libraries of the University of Missouri, the University of Wisconsin, and elsewhere. By and large, these systems represent an effort to incorporate into one card file information as to the location of materials charged out and due date records. A third kind of record can be included in the single file; namely, a record of the identity of the borrower. A variation of this type of system is used by

the Library of Congress in locating talking book machines (14).

Variations on the third class of machine-sorted card circulation systems may be found in the Brooklyn College Library (15), the Detroit Public Library (16) and the Stockton (California) Public Library (17). The principle involved is the same, but in the first case the primary record is a call slip, maintained in numerical sequence and in the others it is a filmed record of the transaction. In both types, one of the most important advantages is that the transaction cards retrieved from returned books may be sorted mechanically into sequence and missing numbers may then be located by matching the returned cards with a perfect deck.

The advantages sought by the application of punched-cards to circulation are speed, accuracy, lowered costs and rapid turnover of stock. All of these gains may have been realized in some libraries: seriously lacking from the literature, however, is evidence that punched-card methods are more efficient than (or even as efficient as) carefully planned manual, photographic or sound-recording methods.

Acquisitions and Accounting

Applications of machine-sorted cards is acquisitions and accounting routines represent, for the most part, nothing distinctive from ordinary business procedures adapted to those physical things peculiar to libraries. A case study has been made of the procedures used in the Columbia University Libraries for accounting (18) and somewhat more brief descriptions of other systems have been written for systems in the Library of Congress (19), the Milwaukee Public Library (20) and others. In England, an allied application is in the maintenance of the stock record (21).

At least in theory, the punched card used as an order or accounting device could become the shelf-list card and/or might be used to select lists of materials by subject. A primary advantage of the card is that it is reusable; that is, its primary cost may be divided by

the number of times it is used. Unfortunately, no good description of the actual use of order cards in this manner could be found. Perhaps the cost advantage is more apparent than real when one begins to add the necessary information (classification number, for example) to the individual card.

There are other theoretical advantages to the use of machine-sorted cards to acquisitions routines. For example, sorting an outstanding order file for materials ordered but not received can be a time-consuming and often frustrating piece of work. Such a job could be accomplished easily and automatically if the order file were on machine-sorted cards and if each included the date of ordering. This and other uses are suggested by Parker (22). Exploitation of these possibilities takes careful planning and a knowledge of the machines which few librarians have. Furthermore, where machine accounting is imposed on the library by a higher authority as a step in centralization of work (for example, accounting), considerations of the central office may preclude design of the cards solely for library purposes.

Analysis of Book Stock and its Use

At least in theory, a device which will enable the librarian to analyze his collection in terms of subject coverage, age of materials, use of materials and other related factors should be of great benefit. Maintaining a "balanced" collection, replacing or removing out-dated items and spotting areas of heavy or light use are important tasks which are at best sporadically accomplished in many libraries. Because of the publicity given to analysis of book stock and use of materials at the Montclair Public Library (23), the impression has gotten abroad that machine-sorted cards present an easy way of accomplishing these tasks.

From conversations with Miss Quigley, this writer believes she would be the first to hold up two cautions: first, no information can be retrieved from cards unless it was recorded on them; and second, planning the system at Montclair was an involved process and at least a measurable amount of time and money has gone into

recording data which has been seldom, if ever, used. However, at least one study of use of a collection using a machine-sorted card method has been carried out independently of the actual circulation system in use (24). Thus, the data necessary to studies of use may either be collected in the process of circulation or by special study while retaining the presumed advantages of the machine-sorted card. Parker (25) and Wight (26) early suggested the use of these cards in circulation studies and in the recording and analysis of library statistics generally. Parker (27) has further suggested that machine-sorted cards may be a useful device in making studies of obsolescence of materials, an area of research in which work is made difficult by the mass of data needed for adequate generalization.

Waugh (28) presented some statistics which illustrate the kinds of information as to who reads what which may be obtained from the Montclair application. At the same time, Waugh indicated, perhaps unconsciously, a danger of that type of analysis if the analyzer is not aware of other important factors; namely, there is the possibility that the selection of materials might be influenced too much by observed use. That is, the poorly served people in a community might become even worse off than before if selection followed use.

Serials Acquisition and Control

The multiplicity of tasks to be performed for serials holdings in reviewing subscription lists, placing of orders, recording of receipts, and charging to various funds will make even the most tradition-bound humanist look for mechanical help. Many types of equipment, of course, have been applied to traditional methods and some departures in policy have influenced procedures.

Moffit (29) has written perhaps the most complete article descriptive of one system of using machine-sorted cards for financial control of serials acquisitions. He presented these advantages as a result: increased control of prices and lists; simplified review of titles subscribed to by subject; and the feasibility of a local union list of serials. The Milwaukee Public Library

(30) has also used machine-sorted cards in connection with its serials records. No description of the last named system could be located. Parker (31) has suggested that financial records, records of receipt of issues and binding schedules may be maintained for serials on machine-sorted cards.

Preparation of Catalogs

A close (and quite probably argumentative) distinction will be observed here to attempt to discuss the use of machine-sorted cards in the routine of catalog preparation and maintenance as against their use in literature indexing and in the manufacture of lists and catalogs for wide distribution. The latter two types of application will be discussed later.

One might suspect that the catalog of the library of the Department of Education of the International Business Machines Corporation (Endicott, New York) should have been placed on machine-sorted cards. It has (32). There, a basic card was prepared for each item and three other cards were reproduced, forming four files. No further description of this application was located. Taube (33) has mentioned very briefly the use of machine-sorted cards to prepare a subject-authority file and to convert that list to book form.

In this area is found a great rarity; reference to an attempt to apply machine-sorted cards which was not continued. Challons, in a discussion following a paper by Perry and others (34) mentioned briefly an effort to prepare a catalog of some 2,000 textbooks in the Technical Library of the Admiralty Signal Establishment. An effort was made to prepare Duplimat stencils directly from machine-sorted cards using a tabulating machine. For reasons not specified, the operation was termed technically not feasible. Machine-sorted cards have been used to make somewhat unconventional catalogs for maps and picture files (35). The Army map service produced an inventory record for its map collection. Such data as area, scale and language were recorded on machine-sorted cards. The resulting file then could be used as a catalog, selection device or charge record.

Langan's "Film 'n File" system involved the insertion of micro reproduction of pictures in machine-sorted cards and the coding of the subject in the balance of the card. Parker (36) suggested the possibility of using the cards in catalog production as early as 1938. Such applications, at least to produce conventional card catalogs, however, either are not feasible or, at least, the actual effort has seldom been made. However, Pike (37) mentioned briefly the transfer of information from machine-sorted cards to catalog cards in a library in England. As will be discussed later, application of punched cards to production of copy for lists and catalogs in book form is feasible.

Shelf-Listing

Any circulation system which requires the punching of a card to represent each item in the collection provides the basic record from which a shelf list could be made. Specific mention, however, of this type of application is infrequent. The Milwaukee Public Library (38) has done its shelf-listing mechanically and Jones (39) has briefly described a system in operation in Stockport, England for maintaining accession records (presumably this could be a shelf list) on Powers Four Cards. The Stockport method does not furnish a complete record of holdings, though it could be modified to do so. A hybrid application, which combines accounting, cataloging, and inventorying is the project carried out in the Library of Congress for listing and accounting for surplus books for veterans at the end of World War II (40). A complete shelf list on punched cards could be used in several ways: to help determine the value of the book stock for insurance purposes; to provide lists of new accessions; and to assist in analyzing the book collection.

Payroll and Personnel Records

Only one reference to the application of machine-sorted cards to personnel records and payrolls was located in the search made for this study (41). However, in all likelihood many institutions having centralized personnel records surely have applied such methods and

Parker (42) has outlined several possible methods. This may be considered a "normal" use of the cards which may be found in many businesses and will not be studied further for present purposes.

Similarly, billing and inventories of equipment are seldom mentioned in library literature and applications of machine-sorted cards here probably offer less of value than in business and industry.

II. B Literature Searching

> There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers - conclusions which he cannot find time to grasp, much less to remember, as they appear....
>
> Professionally, our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose (43).

This basic problem expressed by Bush has led many people to speculate on and to experiment with different mechanical devices to the end of indexing literature so that it might be searched completely and quickly. By the time serious speculation on the problem began, machine-sorted card equipment was being used in other ways. Because of its seeming flexibility and speed, such machines must have appeared to provide a ready-made solution to the problem.

Machine-sorted cards have attracted attention from both admirers and detractors. On the one hand (44), their advantages are described:

> 1. No necessity to establish a methodical coherent system of arrangement. It is sufficient if individual characteristics are identified with consecutive code numbers according to an alphabetical index or other chart. This means that the work

of coding can therefore be completely adapted to the particular task.

2. No necessity to keep the punched cards in any particular order in the file. The cards which are picked out can be put back anywhere in the file.

3. Rapid sorting, even when there are very large numbers of cards.

Fairthorne (45), on the other hand, states that "The disadvantages of punched cards are consequences of their needing but insensitive and coarse means of discrimination, and of being permanent. The power and general cost of handling a large volume of information at even moderate speeds can be very large, and increases as the cube of the speed...." Further, "The arithmetic interpretations of punched-hole positions and subsequent computational operations are not only inconvenient, but also unnecessary, in library work."

At least as early as 1936, the possibility of the application of machine-sorted cards to literature control had occurred to Mayor (46). Since that time, a considerable amount of actual experimentation has been carried out and a good deal of constructive speculation has been published. The literature is scattered through library periodicals, scientific journals, house organs, books, pamphlets and the increasing number of documentation journals. The effort necessary to pull even a fair portion of it together is enough to reinforce the view that mechanical control of literature is highly desirable.

Mechanical literature searching is not universally accepted as the ideal method of preparing for an investigation. Burchard (47), while demonstrating the vast problem which faces the scientists in a literature search, makes the point that simplifying the searching process may result in the loss of "accidental stimuli," which have sometimes been most productive of ideas.

In relation to scope of coverage, at least two major types of machine-sorted card applications to literature control have gone on simultaneously. On the one

hand is the mechanized documentation center, and on the other is the relative narrow and restricted application. As to methods of using machine-sorted cards, at least three major variations have been attempted. In the first case, the effort has been to record all information pertaining to each piece of literature on one card, using trailer cards where necessary; in the second type, the "unit-card" concept has been developed in which each article, report or item of interest is identified by, to use an actual example, raw material, process, product and properties. Four cards are then prepared for each bibliographic item or, if necessary, part of a bibliographic item, and maintained in separate files. In the third type, experimental data are recorded directly on cards. This last, though it can be used as an index to scientific investigations, is more a laboratory than library application and will not be treated in this study. Not all restricted applications will fit into one of these three categories. For example, Stoetzer has described a system in which machine-sorted cards are "keyed" to an existing classed catalog (48). However, the three general methods mentioned appear to cover most systems of particular significance.

Luhn (49) has presented a system for sorting machine-sorted cards photoelectrically. In brief, the scanning machine passes the cards lengthwise under a "question card." Light is directed through the question card into a bank of photo cells. When a card in which a pattern of holes complementary to that in the question card passes beneath the question card, light is shut off to the photo cells, and that card is selected. This machine could, of course, be applied to either a documentation center or a limited field. However, no record of its actual application was located.

Documentation Center

Twenty years ago, Delmas (50) suggested the use of machine-sorted cards in a central location. His idea, in brief, was that requests for information could be directed to the center where partial catalogs or selected bibliographies would be made using mechanical methods. Other similar suggestions have been made,

including the sponsorship of such a documentation center by a library association (51). The underlying bases for the documentation center are the inadequacy of traditional methods of bibliographic control and the potential benefits of concentrating time and money.

A documentation service using machine-sorted cards was reported (52) in 1946. In this application, the cards are used as a subject index, are filed by main subjects and carry bibliographic data on the reverse side.

Pietsch, who has been concerned with operating a documentation center using machine-sorted cards, has written an "Evaluation of mechanized documentation" (53). He describes the problems faced by the Gmelin Institute in publishing the Gmelin handbook and the application of IBM cards and equipment to create a literature searching service. Pietsch concludes that "The present state of electronic development of automatic documentation can be expected to move in the direction of electronic storage units" (54). Thus, he feels that machine-sorted cards are not the solution, or will be used incidentally rather than as a primary storage device.

Machine-sorted cards, then, once appeared to provide a solution to literature searching on a broad scale and have been used in a working system in a documentation center. Their speed in standard equipment, however, appears either to be slower than was supposed or to be less than is essential for searching large files. Shaw (55) has discussed the sorting speed, storage capacity and related problems of machine-sorted cards and other devices which may be used for bibliographic searching. While he states no definite conclusions in reference to machine-sorted cards, it is clear that the storage capacity of such cards is substantially less than the other devices which he examined (56). He also points out that the effective speed of a machine-sorted card system is substantially lower than the speed of the individual machines may indicate (57).

Restricted Applications

Applications of machine-sorted cards to specific files drastically restricted in size and scope as compared with the documentation center concept are numerous. Thorough investigation beyond the literature might reveal a sizeable number not yet reported and might yield new notions in methods. Evidence of the fact that these smaller files are susceptible with some degree of success to indexing and searching by machine-sorted cards, and that a variety of methods may be used, is to be found in the publication resulting from the Symposium or Systems for Information Retrieval, Western Reserve University, Cleveland, Ohio in 1957 (58). Eight of those systems will later be described briefly as more or less typical of other similar applications.

One account of a limited application which was later abandoned was found. Ashthorpe (59), reviewing an attempt to cope with rapidly accumulating report literature at the Atomic Energy Research Establishment, Harwell, England, reported the following disadvantages of a machine-sorted card system as reasons for reverting to an orthodox card index:

1. Especially where multiple sorts were required, the machines were too slow.
2. Machine sorting did not eliminate enough cards.
3. Searching on a "fine" UDC number sometimes meant resorting on a broader number.
4. Wear on cards impaired their usefulness and caused delays.
5. Use of the system was such that the machines stood idle much of the time.
6. No more than one search could be carried on at a given time.

Ashthorpe described the advantages of an orthodox card index over the machine system as:

1. Several searches can be made simultaneously.

2. "Snap" questions are more easily answered.
3. The searcher can adopt fresh approaches during the search.
4. Wear and tear on the file is less - or at least less critical.

It is possible that at least some of the problems Ashthorpe encountered are a result of using the Universal Decimal Classification.

In each of the cases reported in Information Systems in Documentation, an indexing or classification system was designed or adapted for the specific field to be covered. However, coding is not a subject to be discussed in this review and only the methods of using machine-sorted cards will be discussed below.

Single Card-Multi Field Method

At first sight, machines actuated by punched cards appear to operate with great speed. The speed quoted, for example, for the IBM Sorter is 21,000 cards per hour. High-speed sorters reach approximately 60,000 cards per hour. If only one column must be sorted, and if nothing remains to be done after sorting, those speeds are correct. However, literature indexing systems typically require more than a single column; indeed, they frequently require that several multi-column fields be sorted. In addition, cards must be brought to the sorter and taken away after the sorting and cards sorted out of a file must be refiled. Thus, the effective speed of conventional punched-card machinery must be calculated separately for each application and will be substantially below the advertised speeds of some of the machines.

One type of solution to this problem of speed in sorting is represented by a group-selecting device (60). At this stage in the development of mechanical documentation, such devices, while useful, are rudimentary. A more flexible, though undoubtedly more expensive, tool is the machine which will identify a code designation and select the appropriate cards no matter in what part of the cards the code is punched. Garfield (61), in des-

cribing the IBM 101 Statistical Punched Card Machine, states that the device can accept a fluid coding system and that it can make several searches simultaneously, thus at least pointing toward an increase in effective speed. It is interesting to note that, at the same time, Garfield pointed out the advantage of pre-arrangement of cards so as to eliminate some parts of a search.

Unit-Card Approach

Whaley (62) distinguished between "scanning" systems and "collating" systems. In the former, the effort, as described above, is to use one card for each document and to punch into the card several subjects. Then the entire field is sorted for any one of the subjects or, in the IBM 101, several separate sorts may go on at once. The collating system uses, in contrast, the unit card approach, and several cards are made for each article or item of information and maintained in separate files or in an ordered file such that blocks of cards represent definite subjects.

Peakes (63) has described an internal report indexing system which began as a scanning system but which was converted to a collating system after one year because:

> Our attempts to search such a file [a scanning system] revealed that all the cards of the index must be passed through the search machine for every inquiry. For the type of question which we wished to answer, as this kind of index grew, it was placed at an ever-increasing disadvantage as to both elapsed time for an inquiry and the rate at which inquiries could be answered. Also, we were concerned about the possibility of cards wearing out and creating a severe card-replacement problem.

In the revised system, each "unit of indexable information" is tied to a serial number representing the report in which it is found and is coded for specific subject and is further identified as pertaining to one of four broad concepts; raw material, process, product, or

test on the product. A question involving the results of a particular process on a particular raw material is then answered by selecting the pertinent process cards and raw material cards and merging the two decks into a single sequence of serial number. Then, where a process card and raw material card have the same serial number, they will appear together, indicating that the report carrying that number contains some information required for answering the question.

Whaley (64) has described a collating system of document indexing which involves creation of a term index, the assignment of "roles" to terms and the punching of the "address" (i. e. specific location within a report) of each "structerm" into a card and the filing of cards by term number. Role is an indication of the use of a term (as, for example, "Estimate or determine the cost of"). Structerm is defined as a term with a role assigned. The use of structerms represents an effort, reportedly successful, to reduce the number of references resulting from the sorting process. A card is punched for each structerm in each internal report and then the addresses are consolidated on one card for each structerm in a particular report. The consolidation of addresses makes possible a reduction in the number of cards to be retained and sorted. In addition, Whaley proposes to limit the size of the file by removing cards representing old reports and sorting those cards for occasional reference.

Earlier reference was made to several articles on machine-sorted cards in literature searching resulting from the Western Reserve University Symposium, 1957. Following is a brief review of six of them not covered so far.

McCafferty (65) presented the method developed at the Watertown Arsenal to obtain access to literature relating to ordnance. Using a classification system devised by ordnance personnel at Watertown, items of literature were abstracted and classified, punched cards were prepared so that articles or reports could be identified by serial number, source, publication date and subject. Subject and numerical indexes to the resulting

file were created. Bibliographic information and an abstract are contained on each card. It appears that the primary reason for constructing this file might not have existed had the Arsenal's library been adequately organized.

Hayne and Turim (66) described an information retrieval system limited to one chemical compound (chlorpromazine) but which encompasses a wide variety of subjects. Using a modification of the coding system Luhn developed for use with the IBM photoelectric scanning machine (mentioned above), but adapted for the IBM 101, a system was developed in which "...there is no practical limit on the number of factors which may be specified in a search, nor on the manner in which they may be logically combined" (67). The system is not described in detail in the article at hand.

Livingstone and Welt (68) presented a rather detailed description of an experiment in coding of data concerning the relationships of chemical structure to biological activity. The system depends essentially on the matching of files of cards on chemical compounds with files of cards on biological responses. Coded data abstracted from many sources are punched into the card in such a manner that many combinations of factors can serve as the basis for sorting. The Weil and Hildenbrand (69) article is descriptive of a project to abstract literature on fuel and lubricant additives and to create an index to the abstracts on machine-sorted cards. The following files are maintained. Subject and author files of machine-sorted cards, a file of abstracts typed on vellum, a file of handwritten abstracts of both pertinent and non-pertinent materials, and one copy each of abstracts duplicated from the vellum slips is filed by patent or accession number. In addition, a code card is filed behind each abstract. The machine-sorted cards are arranged in the file by a rough compound classification and a color code is used in the file of subject machine-sorted cards. A file of trade names has also been developed. The machine-sorted cards to be sorted may be selected by the file by compound and the search so limited. MacKinnon, Leary and Levinson (70) reported on an experiment in using machine-sorted

cards to legal research processes. Essentially, the experiment involved classifying the Illinois divorce statute and using it as a weeder code. The basic theory is that the Illinois lawyer concerned with a problem on divorce involving the statutes of another state will want only how the laws of the two states differ. Then, the divorce statutes of all other states were similarly classified and points of difference noted. A machine search will reveal those points of difference and the searcher is thus warned to consult the actual statutes.

A modification of classification designed for use with marginal punched cards to a machine-sorted card system, and the operation of that system is the subject of a report by Weil and Clapp (71). The American Society of Metals and the Special Libraries Association metallurgical literature classification is divided into "orders" (first-order, second-order, etc.) by the extent of detail in each concept considered. In the system in question, two fields on the machine-sorted cards are used for recording concepts and concept designations are superimposed on one another in those fields. Work cards, serially numbered, are prepared by the indexer, and are filed by number. Search cards (that is, machine-sorted cards) are punched with bibliographic information, the work card serial number and the concepts indexed. At least two search cards are required for each work card and a set of two search cards is required for every different first-order category in order to eliminate a first-order search of the entire file.

It is perhaps unfair, in view of the fact that this system had been in existence less than a year at the time of reporting, to emphasize that it had then been used "...mostly to locate specific economic figures such as company expansions, metal shipments, or production figures" (72). Or, in short, precisely the kind of information which should be obtainable through standard library tools and techniques.

This discussion of uses of machine-sorted cards is not complete in that it does not review all applications of the cards to literature searching and indexing.

It does, however, indicate the major lines of development from the standpoint of methods. A further and more detailed review could not be accomplished without detailed discussion of coding methods, a subject which is outside the scope of this study.

II. C Extension of Standard Tools

Machine-sorted cards have been applied to the manufacture of long-accepted bibliographic devices such as indexes, catalogs and book lists. In general, these applications have been aimed either at making a central library's holdings accessible to its branches or to other libraries or individuals. The specific use of machine-sorted cards in this type of application is to produce copy from which the catalog or list may be duplicated. Dewey (73) has summarized current applications of this type and has described some methods now being used. It is important to bear in mind in reviewing these systems that the machine-sorted cards and equipment do not produce catalogs; they produce only copy which must then be duplicated. Furthermore, one step in the process which is sometimes overlooked is the completion of a document of some type from which the cards are punched by a key-punch operator. That is, while it would be possible to work directly from the item being cataloged to the punched card, there normally is a step in between.

Several applications of machine-sorted cards to library-related tools were reported immediately following World War II. The Library of Congress listed textbooks used in the ASTP and V12 programs and distributed the lists to colleges in preparation for the anticipated need by returning veterans (74). Arnhym (75) reported on the use of machine-sorted cards for organizing, processing and distributing technical and intelligence information. No description in detail of these processes was located in the literature.

Currently a project is under way at the National Library of Medicine (under a grant by the Council on Library Resources, Inc.) to combine tape-operated

typewriters, International Business Machines and the Listomatic Camera in producing the Current List of Medical Literature (76). While the project is directed toward production of the Current List, it may have side benefits rather similar to those activities of a mechanized documentation center. For example, it may be possible, once the envisioned files of cards are in existence, to produce annual or cumulated bibliographies on broad medical subjects. The Current List of Medical Literature presents some difficulties to the searcher in that complete bibliographic citations are not given under subject headings and the searcher must continually turn from the listings under subjects to a list of articles indexed. In the proposed method, the bibliographic citation will be typed, by tape-operated typewriter, on a machine-sorted card, using only the upper and central portion of the card. The balance of the card may then be used to code the various subjects under which the article will appear. Sorted into sequence by author or by subject, the card will be passed through the camera to produce a film which will serve as copy from which the Current List will be produced. Up to three typed lines of "copy" may be put on each card, thus reducing the number of trailer cards as compared with the method used for New Serials Titles, to be mentioned later.

A proposal to maintain a union catalog of serials at the Library of Congress (77) and to publish it based on a machine-sort card system has been made but has as yet not been put into effect. However, the publication of information on serials titles newly received by a relatively large group of libraries has been carried through by the Library of Congress (78). In contrast to the proposed system being developed by the National Library of Medicine, the New Serial Titles method involves using one card for each line in a given entry. Copy for reproduction is then produced by use of a tabulating machine. The main entry for each serial to be listed requires at least one card (and commonly more than one). The entry cards are then followed by a card for each library for which holdings are indicated.

County libraries have in many cases operated with

many branches and stations spread over wide areas. Typically, those libraries have had to operate on tax bases which are small, as compared to city libraries and in regard to the number of people served. One rather common technique used in those circumstances is to provide the branch or station with a rather small basic collection and then to move the newer materials from one location to another. In many cases, branches and stations have not had complete catalogs but have used shelf lists or other devices. Requests from the branch or station to the central library are then made without knowledge of what the main collection contains, and sometimes without complete knowledge of what material is in the branch itself.

In the King County application (79), machine-sorted cards are used as a locator file and to produce copy for book catalogs of the material in each branch. The branch catalogs each consist of an author list, a title list and a subject list, the latter divided into adult and juvenile titles. The subjects used in the branch catalogs are not identical with those used in the main library's card catalog, and a "key" relating the two systems of subject headings is supplied to each branch. The punched cards are also used to imprint book cards and date due slips with the required information.

The Los Angeles County Public Library has used a different approach to the same basic problem. MacQuarrie (80) described that system and stated its benefits as:

1. The library's entire holdings are open to inspection at any service point in the system.

2. The existence of the book catalog makes it unnecessary to file and pull cards for books as they are moved from one branch to another.

3. Requests for material made by branches to the central library are accurate.

4. Gaps in the book collection stand out more clearly in book form than on cards.

5. The book catalogs result in more effective use of branch collections, especially since many branches did not have catalogs of even their own collections prior to the introduction of this system.

These systems, of course, require that a file of punched cards representing the book collection be created and maintained. They also require some system for keeping the book catalogs up to date. In the case of the Los Angeles County Public Library system, supplements are issued monthly and supplements are interfiled in the bound volumes quarterly (81).

The Columbia River Regional Library, Wenatchee, Washington, has used the Los Angeles County system (82).

These systems are not new in the sense of having created a new library tool. Distribution of a complete catalog of the library's holdings does represent a considerable departure from existing county and regional library methods in the areas mentioned. Of course, other methods of manufacturing those catalogs could have been developed.

The New York State Library, using machine-sorted cards to produce copy has published a check-list of its holdings in the social sciences (83). The aim, of course, is to make the materials held by the State Library more widely known and used.

III. Some Problems

Machine-sorted cards have been applied to three broad areas in librarianship: first, to routine, repetitive operations; second, to literature searching; and third, to the production of copy from which lists and bibliographies are produced. It scarcely needs to be pointed out that all of these operations can be performed by other methods; indeed, the application of machine-sorted cards represents the unusual rather than the standard operation.

To be of significant value, then, the machine-sorted card system must either be less expensive than other methods or it must be capable of producing some end result not obtainable by other methods. The information available in the literature does not demonstrate that machine-sorted cards, applied to library procedures, makes possible the achievement of either goal.

In the first type of application (routine, repetitive operations), there appears to have been an assumption that library records or files are sufficiently similar to business and industrial records that at least some of the advertised benefits will accrue to the library which uses machine-sorted cards. But are they? A circulation file may seem similar to an inventory record of parts. However, the virtue of the circulation card is that it distinguishes one bibliographic unit from all others, while much of the virtue of the inventory card is that it shows the similarity or exact identity of springs or bols or whatever it represents. Thus, almost automatically, one may presume that most cards in a circulation will be used infrequently as compared to most cards in an inventory file.

A common argument in favor of machine-sorted card systems is that they make available information which is not obtainable (or readily so) with conventional methods. For example, Moffitt (84), in describing the use of machine-sorted cards in financial control of serials subscriptions, lists several such benefits. The first of those mentioned is the provision of an annual list of serials to the administrative offices of the library. These lists are "...useful in answering many of the questions concerning serials without the necessity of referring to the serials unit for information" (85). Evaluation of such additional benefits in terms of cost is universally lacking. Furthermore, in this particular instance, it is not clear as to why the serials unit should not be expected to supply information whenever it is asked to do so, and to supply more up-to-date information than could be had from an annual list.

In the second type of application (literature searching), some serious questions are not covered at all in

the literature, yet surely they must have occurred to persons who have experimented with machine-sorted card systems. For example, one virtue of the standard card catalog and of the printed bibliography or index is that it has "integrity" at all times. That is, entries are not removed from them, or if cards are removed from a card catalog, temporary entries show at least the existence of the full entry. This virtue is not to be found in machine-sorted card files. In fact, one of the advantages of the latter type of file is that things can be removed from it. But what happens when a search must be made when another one is going on or has just been completed? Must one wait until items removed have been interfiled again? Or, does one conduct several separate searches of the various files, adding one or more files depending on how many have been created by previous searchers, the results as yet unfiled? and, if one waits, how much time does it take? Again what effect have the processes of coding, machine operation and related processes on the training and, consequently, the salary scales and availability of personnel?

Neither of the two basic requirements noted above (that of cost savings or unique end result) has been documented as a virtue of machine-sorted cards as applied to literature searching. The absence of careful cost data will be commented upon later, but the impression of a group of foreign librarians is of interest here (86):

> The present use of IBM cards for selecting information using commercially available machines has not achieved anything which cannot be achieved by traditional methods, although it may in certain cases be cheaper than manual methods, but adequate costing data were not obtained.

Thus, the basic questions have not been answered, and other serious questions have been raised but also left unanswered.

In the third area (production of copy for catalogs and lists), the case is also unclear. Many devices -- standard type-setting machines, photo-offset machines

-- and various methods (87) exist for producing copy to be duplicated. The literature does not reveal a single attempt to detail the present cost of producing copy for a catalog or list as compared to the cost where other devices and methods are used. Neither does it show careful analysis of the end product, including factors such as storage capacity per page of copy.

IV. Cost Data

A few of the articles reviewed for this project contain some information about the costs of the system described. In at least one case (88) there is evidence to indicate that more or less complete cost information might be available for inspection. In another case (89), a cost study is in progress. Fragmentary data is recorded in a few other cases, and it may be that in some of those instances cost data might be obtainable.

To be reliable, cost data must be sufficiently detailed and collected with sufficient care that one could reconstruct a procedure and rely upon being able to predict its performance and cost with fair accuracy. So defined, cost data does not exist for applications of machine-sorted punched cards to libraries in any instance. Or, if it does, it is not obtainable in the literature.

V. Suggestions for Research

Future research in the application of machine-sorted cards to libraries will be most fruitful if it can yield some detailed information on the two key questions formulated earlier:

1. Can things be accomplished through using machine-sorted cards which cannot be accomplished through other methods?

2. What are the costs of machine-sorted card systems as compared to other methods?

Following the general outline of this paper, other important questions, not answered in the literature, are:

I. Routine, repetitive operations

A. What operations in libraries are best suited to the application of machine-sorted cards?

B. Can some principles or general rules be developed by which to estimate success in advance of actual installation of machine-sorted card systems?

C. Can data be developed which will show with some precision the effect of machine-sorted cards and related equipment on the total process? That is, can the benefits of machine-sorted cards be isolated from the other parts of total systems so that it may be possible to combine various devices or methods into a "best" system?

II. Literature Searching

A. What is the effect of searching a file upon succeeding searches in terms of adequacy of coverage, time required and the opportunity for error?

B. What demands do machine-sorted card systems make, from the encoding process to the final product, upon the personnel who operate the system in terms of training, supervision and availability? What do these demands imply for the library in personnel recruiting and instruction?

III. Preparation of copy for lists and catalogs

Because many tools and methods for preparing copy for reproduction exist, the most important question to be resolved here will be that of cost as compared to other methods. Information on speed of production of

copy, ease of preparing cumulations and storage capacity of the resulting page will also be important, but, again, on a comparative basis.

References

1. Casey, R. S., et al, eds. Punched cards; their applications to science and industry (2d ed., New York, Reinhold, 1959), pp. 638-671.
2. Parker, R. H. Punched-card method in circulation work. Library journal 61:903-905 (1936).
3. Wight, E. A. Methods and techniques of library surveys. In: L. R. Wilson, ed., Library trends (Chicago, The University Press, 1937), pp. 344-360.
4. Wight, E. A. Public library finance and accounting. (Chicago, American library association, 1943), pp. 88-92.
5. Maxfield, D. K. Library punched-card procedures. Library journal 71:902-5, 911 (1946).
6. Blasingame, R. U. Applications of International Business Machines in libraries. (M. S. essay, Columbia University, 1950), p. 15.
7. Pike, J. R. A future for mechanization? Library association record 56:47-49 (1954).
8. Berry, M. M. Applications of punched cards to library routines. In: Casey, Punched cards ..., pp. 279-302.
9. Gull, C. D. A summary of applications of punched cards as they affect special libraries. Special libraries 36:208-212 (1947).
10. Parker, R. H. Library applications of punched cards: A description of mechanical systems. (Chicago, American library association, 1952) 80 p.
11. Quigley, M. Library facts from International Business Machines cards. Library journal 66:1065-1067 (1941).
12. Callander, T. E. Punched card systems in the public library. In: Library association. Conference, 1947. Brighton. Papers and summaries

of discussions at the Brighton conference. pp. 23-28.
13. Duer, M. D. and Lewis, C. S. How we use IBM. Library journal 78:1288-1289 (1953).
14. Berry, op. cit., p. 294-295.
15. Blasingame, op. cit., p. 44-47.
16. Parker, op. cit., p. 26.
17. Klausner, M. IBM circulation control. Library journal 77:2165-2168 (1952).
18. Blasingame, op. cit., p. 40-44.
19. Keller, A. H. Book records on punched cards. Library journal 71:1785-1786 (1946).
20. Baatz, W. H. and Maurer, E. H. Machines at work. Library journal 78:1277-1281 (1953).
21. Pike, J. R., op. cit.
22. Parker, op. cit., p. 13-14.
23. Quigley, M. Ten years of IBM. Library journal 77:1152-1157 (1952).
24. Morse, W. F. and Erickson, E. W. A report of procedures used in a circulation survey of a public library. University of Illinois library school. Occasional papers no. 29, 1952.
25. Parker, R. H. Punched-card method in circulation work. Library journal 61:903-905 (1936).
26. Wight, E. A. Methods and techniques of library surveys. In: L. R. Wilson, ed., Library trends (Chicago, American library association, 1937), p. 344-360.
27. Parker, R. H. Library applications of punched cards: a description of mechanical systems. (Chicago, American library association, 1952), p. 19-20.
28. Waugh, D. What they are reading. Publishers' weekly 146:492-493 (1944).
29. Moffit, A. Punched-card records in serials acquisition. College and research libraries 7:10-13 (1946).
30. Berry, op. cit., p. 284.
31. Parker, op. cit., p. 14-15.
32. Gull, op. cit.
33. Taube, M. New tools for the control and use of research materials. American philosophical society proceedings 93:248-252 (1949).
34. Perry, J. W., et al. The use of punched cards in

American libraries. In: Library association. Conference, 1947. Brighton. Papers and summaries of discussions at the Brighton conference. pp. 40-50; Discussion, pp. 50-54.
35. Gates, M. L., et al. Punched cards for library records. Library journal 71:1783-1786 (1946).
36. Parker, R. H. Mechanical aids in college and university libraries. American library association proceedings, 1938, pp. 817-819.
37. Pike, op. cit.
38. Baatz and Maurer, op. cit.
39. Jones, E. G. Mechanization of accessions records. Library association record 56:214-215 (1954).
40. Gull, op. cit.
41. Baatz and Maurer, op. cit.
42. Parker, R. H. Library applications of punched cards: a description of mechanical systems. (Chicago, American library association, 1952), p. 45-46.
43. Bush, V. As we may think. Atlantic monthly 176: 101-108 (1945), p. 101.
44. International federation for documentation. Manual on document reproduction and selection. 746.11 E2.
45. Fairthorne, R. A. Information theory and clerical systems. Journal of 9:101-116 (1953), p. 107-108.
46. Mayor, Y. Applying statistical machines to bibliographical investigations. Chimie et industrie 35:458-466 (1936).
47. Burchard, J. E. Waterloo of science. Review of documentation 16:94-97 (1949).
48. Stoetzer, W. Die anwendung der IBM - lochkarten in sonderbereichen der organischen chemie. Nachrichten für dokumentation 3:189-192 (1952).
49. Luhn, H. P. The IBM electronic searching system. (Poughkeepsie, N. Y., The IBM Corp., 1952) 11 p. and appendices. Planographed.
50. Delmas, J. Les machines a cartes perforees; possibilitees d'application dans un centre de documentation, notament en vue d'etablir des catalogues. International federation for documentation transactions 14:115-116 (1938).
51. Fitzgerald, F. E. Special librarian and technical

progress. Special libraries 37:35-40 (1946).
52. Roches, C. La mecanographie au service de la documentation. L'industrie textile 63:107-108 (1946).
53. Pietsch, E.H.E. Evaluation of mechanized documentation at the Gmelin institute. In Casey, Punched cards ..., pp. 571-618.
54. Ibid., p. 612.
55. Shaw, R. R. Management, machines and the bibliographic problems of the twentieth century. In: Shera, J. H. and Egan, M. E., eds. Bibliographic organization (Chicago, The University press, 1951), pp. 200-225.
56. Ibid., p. 208.
57. Ibid., p. 210-211.
58. Shera, J. H., et al., eds. Information systems in documentation (Advances in documentation and library science, vol. II. New York, Interscience, 1957). pp. 229-403.
59. Ashthorpe, H. D. Punched-card indexing experiment at the library of the atomic energy research establishment, Harwell. Aslib proceedings 4:101-104 (1952).
60. For example, see Weil, B. H., and Hildenbrand, G. Information on fuel and lubricant additives from machine-sorted punched cards. In: Shera, Information systems, pp. 271-293.
61. Garfield, E. Preliminary report on the mechanized analysis of information by use of the 101 statistical punched card machine. American documentation 5:7-12 (1954).
62. Whaley, F. R. A deep index for internal technical reports. In: Shera, Information systems ..., pp. 352-383.
63. Peakes, G. L. Experience with the unit card system for report indexing. In: Shera, Information systems ..., pp. 306-327.
64. Whaley, op. cit.
65. McCafferty, J. Retrieval of information on armor and kinetic-energy amor-defeating ammunition. In: Shera, Information systems ..., 229-239.
66. Hayne, R. L. and Turim, F. Machine retrieval of pharmacologic data. In: Shera, Information systems ..., pp. 240-249.

67. Ibid., p. 247.
68. Livingstone, G. A. and Welt, I. D. Chemical structures and responses of organisms to applied chemicals. In: Shera, Information systems ..., pp. 250-270.
69. Weil, B. H. and Hildenbrand, B. Information on fuel and lubricant additives from machine-sorted punched cards. In: Shera, Information systems ..., pp. 271-293.
70. MacKinnon, F. B., et al. An analysis of an experimental adaptation of punched cards and mechanical retrieval to legal research and indexing of state statutes, codes and session laws. In: Shera, Information systems ..., pp. 294-305.
71. Weil, B. H. and Clapp, E. A. The adaptation of the ASM - SLA metallurgical literature classification for use in a machine-sorted punched card system. In: Shera, Information systems ..., pp. 328-351.
72. Ibid., p. 346.
73. Dewey, H. Punched card catalogs - theory and technique. American documentation 10:36-50 (1959).
74. Library of Congress reports speedy distribution of textbooks. Publishers' weekly 150:2346 (1946).
75. Arnhym, A. A. Aviation engineer, take notes. Aero digest 53:25, 167, 186 (1946).
76. Information based on conversations with Col. F. B. Rogers and Mr. Seymour Taine, National library of medicine.
77. Keller, A. H. A plan for the development of a union catalog of serials by punched card methods. (Washington, D. C., Library of Congress, 1950) 15 p. Planographed.
78. New serial titles. Washington, D. C., Library of Congress, 1953-
79. Martin, A. F. IBM catalogs for the King County public library, Washington State. (Master's thesis, Western Reserve University school of library science, 1953). 42 p.
80. MacQuarrie, C. IBM book catalog. Library journal 82:630-634 (1957).
81. Letter from John D. Hendersonk, Los Angeles County librarian, January 5, 1959.

82. Cutler, D. The Columbia River regional library demonstration. Library resources and technical services 2:181-182 (1958).
83. Checklist of books and pamphlets in the social sciences. (Albany, The New York State library, 1956). 142 p.
84. Moffitt, op. cit.
85. Ibid., p. 12.
86. Documentation techniques in the USA; selection, reproduction, dissemination. Report of a group of European experts. (Paris, Organization for European economic co-operation, 1951).
87. Gull, C. D. The cumulative catalog technique at the Library of Congress. American documentation 2:131-141 (1951), for one example.
88. Klausner, op. cit.
89. Letter from John D. Henderson, Los Angeles County librarian, January 5, 1959.

Volume Four Part Four

ELECTRONIC SEARCHING

by

Gerald Jahoda

I. Introduction

A. Scope of the Survey

The literature on the use of electronic machines for information handling and closely related operations is reviewed in this report. A differentiation has been made by Mooers between two types of systems: an information retrieval system, i. e. a system which comes up with serial numbers or bibliographic citations of documents in which the answer to a question might be found, and a question answering system, also called a data file, which yields the answer directly--e. g. a chemical, a property of a substance. This differentiation we consider useful and it forms the basis of arranging these two types of information systems in this report. Every other type of application has been grouped under the commonly used miscellaneous heading. In this group machine applications which might assist the librarian in the intellectual task of preparing his subject authority list are included, as are applications aimed at eventually relieving the librarian from the intellectual tasks of abstracting and indexing.

The literature on language translation by machine has been omitted not because it is not pertinent to the overall problem of information retrieval (it is) but because of the vast literature in this field and this reviewer's lack of special knowledge of it. Business applications of computers for inventory and various accounting operations and scientific and engineering calculations by computers have been omitted as being outside the scope of this review.

B. Machine Operations and Characteristics

Modern data processing machines combine three

major techniques which are described by Alexander:

1. Transmitting encoded information as electrical signals, in the sense that one transmits a telegram consisting of both numerical and alphabetic texts.
2. Means for storing such encoded information in a form that permits the information to be recovered selectively as electrical signals when needed.
3. Means for processing encoded information in accordance with the rules of simple arithmetic and the rules for elementary logic (1).

If optical signals are added to electrical signals, then the description also applies to machines which process data on film, namely the Rapid Selector, the Filmorex, and the Minicard systems. Before any use can be made of these internal characteristics of the machines, information which is to be processed--index entries, i.e. index headings and some kind of document identification--has to be translated into a form which is easily handled by the machine. This consists of translating the index headings and the document citation into short hand symbols, generally numbers, called the code. The code is then converted into a machine language, i.e. a language which the machine can understand. This is done by converting the numbers into a series of transparent and opaque patterns for machines which sense optical signals, or by converting the code into holes on punched cards or punched tape. In most computer based systems the code on punched cards and punched tapes is converted into magnetized spots on magnetic tape or magnetic discs.

The machine is now ready to manipulate this data, that is, to perform simple arithmetical and logical operations with the data, if it is instructed exactly what to do. The sets of instructions are called the program. Programming operations vary from machine to machine. In the case of machines which sense optical signals this may be done by inserting a mask which complements the pattern being sought in the reading station. In the case of punched card sorters electrical contacts have to be

made inside the machine by setting dials or switches or by making these connections with a control panel board. A control panel board, also known as a plug board, contains a number of holes into which wires are inserted. These wires make connections with wires inside the machine when the control panel is put into place. In the case of computers, programming is done by sending instructions into the machine on punched cards or punched paper tape or magnetic tape.

The program instructs the machine what to do with the encoded data, i. e. the index entries. The basic instruction is to match a search heading against an index heading. Most machines can match more than one search heading against more than one index heading in a document. The type of operation which a machine can perform is often represented by symbolic logic. An index heading or a part of an index heading is represented by a letter, for the sake of convenience. Thus any 5 index or search headings or parts of index or search headings can be represented by A, B, C, D, E.

A search for a document containing either A or B or C or D or E is called a search for a logical sum.

A search for a document containing A and B and C and D and E is called a search for a logical product.

A search for a document containing A but not B is called a search for a logical difference.

The programmed machine scans the encoded record and recognizes items which meet search specifications. Once the search has been conducted the machine has to produce the results in a form understandable to human beings. In some machines this is done by separating the units of information; e. g. punched cards which identify pertinent documents are physically separated from other punched cards. In other machines this is done by printing out the documents by identifying serial number. In still other machines the search results are copies of the abstracts of pertinent documents.

Machines differ in speed, versatility, and cost.

The difference in versatility in terms of potential capabilities for information retrieval has been analyzed by Perry, Kent, and Berry. They state that subject content of documents may be significantly distinguished on the basis of the following two different types of characteristics:

> Type I Characteristics: Spatio-temporal entities (substances, devices, organisms, persons, etc.), attributes, abstract concepts, processes, locations, and conditions involved;
>
> Type II Characteristics: Relationships involved between the entities, attributes, concepts, processes, locations, and conditions.
>
> In ordinary writing, the first type of characteristics is usually designated by nouns, adjectives, verbs, and adverbs. Relationships--our Type II characteristics--on the other hand, are denoted by the phrasing of sentences, by such grammatical devices as endings and other affixes or connectives such as prepositions (2).

Different types of machines are analyzed by these authors in terms of allowing searches of various complexity with Type I characteristics (descriptors) and Type II characteristics (relationships among descriptors).

Type I devices are the simplest possible devices which would record only one characteristic for each of the documents and would direct searching operations to the characteristic or characteristics that correspond to the item or items desired. An example of this type of device is the National Bureau of Standards Microimage Selector, in which a document is merely identified by its location.

A Type II device is illustrated by the Rapid Selector, as demonstrated several years ago at the United States Department of Agriculture (3). Each unit of information on a frame of microfilm was identified by 6 index entries (more than one frame could be used per item of information). Only one entry could be searched at any one time. This application of the Rapid Selector

could be construed as a speeded-up search in a conventional index.

Type III devices can be used to record a multiplicity of criteria (descriptors) in fixed zones or fields on the recording medium. The search can be made for one or more criteria in fixed zones. Standard IBM or Remington Rand sorters are examples of Type III devices. In the case of single column sorters, searches for combinations of--for example-4 descriptors have to be made by sending the cards through the machine 4 times. In the case of multiple column sorters a similar search might be completed in one sort through the machine. An important limitation of these machines is their inability to search for Type II characteristics, namely relationships among descriptors.

Type IV devices are row-by-row punched card scanners as exemplified by the Luhn scanner. In addition to being able to do everything that Type III devices can do, Type IV devices can treat consecutive IBM cards as one continuous unit. Type II characteristics can also be indexed and searched. The combination of these factors makes the Type IV device a much more flexible device for searching.

Perry, Kent, and Berry's Type V device can handle Type II characteristics in a more sophisticated manner than that ascribed to Type IV devices. For example, the machine can differentiate between "Man bites dog" and "Dog bites man," "Blind Venetian" and "Venetian blind." The machine has also the capability of detecting the beginning and end of sequences of descriptors. Since Perry, Kent and Berry's text on machine literature searching was written, a Type IV device--the ILAS (Interrelated Logic Accumulating Scanner)--has been built which is capable of differentiating such relationships. Consequently the difference between Type IV and Type V criteria is no longer as sharp.

Perry, Kent and Berry's Type V criteria is a refinement of their Type V device. The machine can "interpret" the meaning of a descriptor by means of a lookup table in its memory before starting on its matching

operation of descriptors and descriptor connections (4). A general purpose computer based system such as the IBM 704 Electronic Data Processing Machine can probably be programmed to include Type VI device criteria.

Machines which can recognize electrical signals can be grouped into 3 categories, which are arranged in increasing order of complexity:

1. Machines which can only handle information entered on one punched card as the largest unit. These machines can recognize combinations of descriptors but cannot recognize stated relationships among descriptors. Conventional column-by-column scanning punched card sorters are examples of this type of machine.

2. Machines which can handle information entered on one or more punched cards as a single unit. These machines can recognize combinations of descriptors and stated relationships among descriptors. This is not because of any additional capacity for performing logical or mathematical operations in the machine but because a code which indicates relationships among descriptors can only be conveniently prepared if more than the 80 columns of a punched card can be used as a continuous unit for any one document. Row-by-row scanning punched card sorters such as the ILAS are an example of this type of machine.

3. Machines which can handle information entered on one or more punched cards (or the equivalent on magnetic tape) as one unit and can store preliminary results of the searches in their memory. These machines can recognize combinations of descriptors, stated relationships among descriptors, and can introduce various refinements in the search. An example of this is the assignment of a numerical value to each descriptor in a search based on its significance for that particular search and the selection of documents which contain descriptors with a sum total of a stated numerical value. Computers, such as the IBM 704 Data Processing Machine, are an example of this type of machine.

We may also list briefly the individual components

or units of computers and Type I and II punched card sorters. A digital computer generally consists of these five parts:

> An input unit where information is received from the outside world in machine language;
>
> A storage or memory unit where information is stored or remembered and ready for use in a large number of locations or registrars;
>
> An arithmetic unit where information is operated on arithmetically or logically;
>
> A control unit, the unit which controls the switches or gates that connect specific registers in the units and thus controls the sequence of operations in the computer;
>
> An output unit where information is returned to the outside world, usually in a form readable by human beings.

These five units are connected so that instructions and information (both in the form of electrical impulses) can flow from one unit to another (5).

Punched card sorters, on the other hand--as exemplified by the IBM 101 Electronic Statistical Machine --are much simpler devices. This particular multiple column sorter consists of several components all incorporated into one unit. These components are:

> The input station, which is the card feeding device, also known as the hopper;
>
> The card reading station, which is equivalent to the computer's arithmetic unit in that it is the part of the machine where information on individual cards is examined;
>
> The instruction station, which in this case is the control panel where wires are connected to the card reading station to perform the desired oper-

ation;

The output station, where information is returned to the outside world either as punched cards sorted in a particular pocket or as a line of print generated by the machine's print-out unit.

The row-by-row scanning punched card sorter, as exemplified by the ILAS, has a separate instruction station which is connected to the input, card reading, and output component. This particular machine does not have a print-out component.

Detailed descriptions of these machines can be found in manufacturers' instruction manuals such as the IBM 101 manual (6) or in general texts on computers such as the books by Berkeley and Wainwright (7) and by Chapin (8).

II. Information Retrieval Systems

A. Types of Indexing Systems

The shopper for an indexing system can make his selection from several packages, the contents of which can be custom blended. He can choose a well-worn and long-tested traditional package, the alphabetic subject index, the alpha-classified index, or the classified index. He can also choose a newer package, some form of coordinate index. Or he can make a reservation for the package that is still a gleam in the manufacturer's eye, the index prepared by machine.

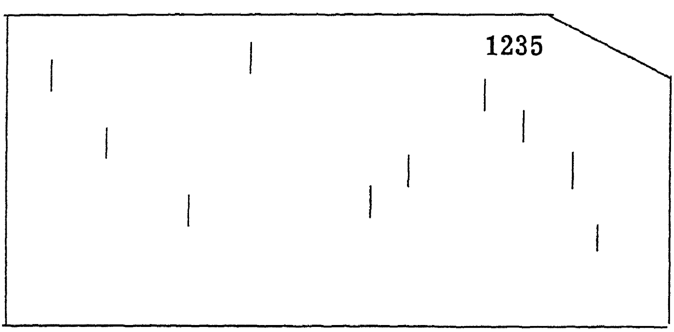
Coordinate index is a generic term for an indexing system which departs from the one-to-one relationship of the index heading to the citation as a unit as it exists in conventional indexing systems. On one physical unit such as an index card, a punched card, or a section of paper or magnetic tape, more than one index heading is generally associated with one document citation, or else more than one document citation is associated with an index heading. This is illustrated by the following unit records of a conventional indexing system and two coordinate indexing systems:

Conventional alphabetic subject index

Jet fuels, thermal stability

1235

Coordinate index, conventional grouping:
index heading on document card



Coordinate index, inverted grouping:
document numbers on index heading card

Jet									
0	1	2	3	4	5	6	7	8	9
120	311	12			1235				689
	411	52							
		62							

In the illustrated examples the serial numbers identify the document. The serial number is the equivalent of an abbreviated citation. The holes on the punched card stand for index headings which are assigned to that particular document.

The physical association of more than one index heading to a document citation or more than one document citation to an index heading permits manipulations of the coordinate index which are not possible with a conventional index. A search can be made for a document which has two or more index headings in common, which has one index heading but not another, or which has one of several specified index headings, by a mechanical matching of numbers or holes. (The critical reader will no doubt say that this is what we are doing all the time when we search a card index under one heading and make our decision of accepting or rejecting an item on the basis of the existence of another heading mentioned in the tracings. This is certainly true; but with the coordinate index one can carry this to greater heights of sophistication and one can do it by machine.)

Greater advantage is taken of this ability to coord-

inate headings by splitting down index headings into smaller but more generic units, usually called descriptors. For example, the index heading "jet fuels, thermal stability" might be broken down into 4 descriptors: thermal, stability, jet, and fuels. Each of these descriptors, either singly or in any combination, can then be used as an access point to the index.

By splitting up the conventional index heading into descriptors, the relationships among the individual parts of the index heading are lost. In some of the coordinate indexing systems this loss is not considered serious. The argument is that information will not be lost but that additional though extraneous documents will be selected. It is easier to separate these extraneous documents manually than to prevent their appearance in the first place. In other coordinate indexing systems the relationship among descriptors in a document (other than the sometimes accidental and misleading relationship of being in the same document) is brought out in the index in order to reduce the yield of extraneous documents in a search.

The third type of index is in its early development stage. It differs from the conventional and coordinate index not so much by its form of entry but by its method of preparation. After the preliminary work of the librarian is completed the machine will prepare the index to the documents, and--given a request in essay form--the machine will translate it into its index language and perform the search.

B. Intellectual and Mechanical Aspects

Information Retrieval

Indexing systems, whether conventional or coordinate, are shaped by two major factors: the stipulated users' requirements and the financial resources for developing and operating the indexing system. These factors are based on the following considerations:

Type of user: research scientist, patent attorney,

administrator, engineer, student or general public; specialist's use of system in his field or in some one else's field;

Size of installation in terms of number of documents: present size, rate of acquisition, expected maximum size;

Subject matter of collection: its homogeneity or heterogeneity;

Other available indices to collection;

Type of anticipated searches: specific, generic, correlative, predominance of any one type of search;

Frequency of searches: per given period, at one time;

Search results: required completeness, required up-to-dateness, required speed of completion, required freedom from non-pertinent references;

Type of use of index: on a self-service basis, through librarian, in multiple locations, in central location;

Users' and management's attitude toward existing index;

Time and resources available for development work on system;

Time and resources available for incorporating backlog material into system to make it useful at an earlier date;

Time and resources available for routine operations of system;

Availability of data processing machines.

The sum total of these variables constitutes the environ-

ment of the system. Just as different variables make up the environment of the system, so the indexing system itself is made up of a number of variables. Calling an index an alphabetic subject index or a coordinate index identifies its genre in a broad way but gives little information about the index itself. Some of the variables in an indexing system will, therefore, now be identified. For the sake of convenience these variables will be grouped into intellectual and mechanical aspects of the index.

Intellectual Aspects

The indexing and searching operations in any given installation and for any system except one in which indexable information is selected by machine involves the following operations: A document selected for inclusion into the system is read for indexable information as specified by a set of rules. The depth of indexing will influence the time spent in reading (or scanning) the document. A time limit for reading any one article might be set or instructions might be given to read only certain parts of the document, such as the summary, the table of contents, the conclusion, or the abstract.

The reading so far has been for indexable information. The translation of this information into index language, again according to a set of instructions, is the next step in the process. New terms are incorporated into the system at this stage, either as indexing terms or as cross-references. The last step is a mechanical one and consists in the actual preparation of the index entries according to the various systems and in the finding of a "parking place" for the indexed document.

For the retrieval operation some of the above steps are reversed. The inquirer poses his question in terms of the index language. This again requires the translation of communicative language into the index language and is often done with the assistance of the indexer or some one else familiar with its operation in order to bring the thoughts expressed in two relatively different languages into coincidence. The index is then

searched under these entries, either manually or mechanically, by the final user of the information or by some intermediate or by machine. The question may be phrased differently a second time if the wrong index headings have been selected or if the inquirer adjusts his search as a result of information obtained through the first search of the index. This process may be repeated several times. The number of times is dependent on several factors, such as necessity for completeness of information, amount of time available, perseverance, and ability to define what is wanted.

A number of variables or decision points are involved before the indexing system can be put into operation.

Selection of documents to be included into system:
The selection might be by form of literature, e.g. all internal reports, or by subject, e.g. all published and unpublished information on a particular group of chemicals.

Depth of index:
This involves the amount of information in documents to be included in the index. To use extremes, index entries might be made from the title only or from every bit of information contained in the entire document.

Point of view of index:
This will depend on the present and anticipated interests of the users. Decisions have to be made whether to index from the author's point of view only, from the present user's point of view, from all possible points of view, or from any point on this spectrum.

Specificity of the individual index entry:
This is the amount of detail included in the individual index heading. Decisions will be based on the size of the installation, the amount of information now available and anticipated on a given subject, the user's interest in a given subject, and the indexer's philosophy on retrieving non-pertinent

information.

Structure of the index heading:

A choice has to be made between words as headings, phrases as headings, and in some coordinate indexing systems arbitrarily defined units which are combined to make up the index heading. An example of the latter is the characterization of a chemical not by its name but by two or more identifying characteristics such as the number of carbon atoms, the type of bonds connecting the elements which make up the chemical, or the functional groups which are present in the chemical. The index may also be made up of several levels of generality as in a classification system.

The arrangement of the index headings:

The arrangement of the list of index headings can be alphabetical, classified, or alpha-classified. In case of a classified arrangement an alphabetic index to the system is necessary to permit access to the information. A variation on the traditional hierarchical classification is the grouping of the terms in a small number of relatively broad categories. The total number of terms is often small so that the terms in each pertinent category can be read whenever an item is indexed or whenever the index is searched.

The arrangement of the constituent parts of the index heading:

This is to a certain extent a factor determined by the type of indexing system. The arrangement will range from a completely ordered set of terms in the case of a facetted classification system to no particular order whatsoever in some alphabetic indexing systems. Coordinate indexing systems consisting of single words only sidestep this particular problem.

Relationships among indexing terms:

The relationships among terms in a conventional indexing system, i.e. an alphabetic subject, alpha-classified, or classification system, are

brought out by word order (blind Venetian vs. venetian blind), prepositions (reaction *of* benzene vs. reaction *in* benzene), and punctuation symbols (Chemistry, analytic, vs. Chemistry-Bibliography). In coordinate indexing systems these relationships are either not brought out (descriptors are merely listed but not related to each other) or are brought out by means of the following devices:

1. The modification of the descriptor to reduce the scope of its meanings: Benzene (reactant) or benzene (solvent);

2. The assignment of an additional code to denote descriptor order or relationships among descriptors: Venetian--1, Blind--2.

Control of vocabulary of indexing terms:

The two extreme cases are a subject authority list for all indexing terms used in the system and the selection of terms from the documents without any control of the indexing vocabulary. The subject authority list, a defined list of indexing terms along with a network of cross-references, is unquestionably the preferred approach. Intellectual and cost problems, however, often preclude its preparation.

Number of indices to the collection:

In addition to any published index available for the collection, more than one index is sometimes prepared. This is particularly desirable when part of the information has to be indexed in greater detail and/or doesn't fit into the overall pattern of the index.

Completeness of search results and amount of extraneous material retrieved along with the pertinent search results:

These two factors are interrelated since most systems cannot be designed to yield all the pertinent material without any non-pertinent material. The desired amount of pertinent material and the tolerated amount of non-pertinent material will

have a bearing on the design of the system. Two types of non-pertinent results exist, only one of which occurs in conventional alphabetical, alpha-classified, and classified systems. The type of non-pertinent result which occurs with conventional and coordinate indexing systems is often referred to as the "noise" of the system. It occurs when the selected document falls within the defined scope of the index heading but is of no interest for that particular search. An example of this would be a search for poodles in an index in which the most specific heading was "dogs." A document on cocker spaniel is legitimate as far as the system is concerned but it is of no use for the search in question.

The second type of non-pertinent search result is called a false drop and it occurs only in coordinate systems. It is due to the interaction of unrelated indexing terms in the same document or the interaction of unrelated parts of the code. This is illustrated by the following example: Two subjects which occur in the same document are:

Property X of Chemical A
Property Y of Chemical B

This is indexed as Property X, Chemical A, Property Y, and Chemical B--four indexing units, called descriptors, which are tied together by means of a common document serial number (the document's identification). The relationship among the 4 descriptors is not specified. Consequently, a search for Chemical A which has Property Y will yield this particular document even though it does not contain the desired information.

Mechanical Aspects

The manner in which information is stored, the way it is searched, and the form of search results obtained depend to a large degree on mechanical aspects of the system.

Order of descriptor (index heading) to document citation (document serial number):

Three choices are possible: the listing of all descriptors for a document under that document number, the listing of all the document numbers which are characterized by a descriptor under that descriptor, and the listing of one descriptor to one document number.

Type of code:

In both conventional and coordinate indexing systems, the index heading, or the descriptor, may be written as words (as in the case of the alphabetical subject index or the manual Uniterm system) or as short hand symbols, called either the notation or the code (as in the case of classification systems or hand and machine sorted punched card systems). The code in a hand or machine sorted punched card or computer based system is translated into a form which is best manipulated or "read" by the machine and which makes most efficient use of the available space. Several types of codes are exemplified:

> Direct code: Each position is assigned a meaning completely independent of meanings assigned to other positions. For example, the position characterized by column 22 row 8 on an IBM card might stand for the descriptor A;
>
> Indirect code: The significance of a given position is dependent upon its combination with another position. An example of this would be the assignment of column 24 row 9 and column 30 row 1 of an IBM card to mean descriptor B.

Much of the coding in machine based systems is of the indirect variety. The example above is called a numerical code. Other varieties make use of combinations of numbers available from 4 positions. For example, any number from 0 to 9 can be obtained from 4 positions assigned the meaning 7, 4, 2, 1 and using a maximum of 2 positions for each number. (0=no punch, 1=1, 2=2, 3=2 and 1...

9=7 and 2.)

A random superimposed code is another type of indirect code. In this code 2 or more numbers are entered in a code field (a number of columns on an IBM card, for instance) in such a way that the numbers partially overlap or are superimposed on each other. This is illustrated in the following 10 column code field into which 3 codes of 4 positions each are entered:

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Codes: 2, 14, 25, 92
17, 38, 42, 63
27, 31, 53, 82

The interaction of individual positions of these codes produces erroneous combinations such as 2, 14, 17, 25 etc. By a judicious assignment of codes, however, this type of erroneous or false combination can be kept to a minimum in the searching operation.

Another differentiation which should be made is between a fixed field code and a free field code. All of the codes illustrated above are fixed field codes in that they refer to a particular

portion or combination of positions on the storage medium. In a free field code a particular pattern of codes rather than a position of codes is specified. An example of a free field code is a code which is characterized by a pattern of punches as represented by punching the first, third, and fifth punch in a row on an IBM card. The search now is not for position 1, 3, and 5 in a specified column of the IBM card but for this pattern found anywhere (or in a restricted number of columns) on a card.

The operation of a given machine will dictate the way in which any of the above codes are entered on the storage medium. In machines which sort punched cards, column by column, e.g. the IBM 101, the code is most frequently entered as standard alpha-numeric Hollerith code. Any digit from 0 - 9 is entered as a punch in one of the 10 positions of an IBM card column (the eleventh and twelfth positions are disregarded when numbers are punched). A letter is punched as 2 positions in the 12 positions of the column, one of which is in the eleventh or twelfth column. In most computer based systems the code is translated into binary digits, that is, into a system where any number is represented by units utilizing the base of 2. This unit represents 1 or 0, translated into electrical impulses in an on or off position. Any number from the more conventionally used system with a base of 10, i. e. the decimal system, can be translated into the binary system. For example, 0=0, 1=1, 2=10, 3=11, 4=100, 5=101, 6=110, 7=111, 8=1,000, 9=1,001, 10=1,010, 11= 1,001. Not all binary codes are translated into the decimal system or make use of the decimal notation. Another notation used with the binary code is the hexadecimal notation which uses a base of 16 instead of the base of 10. The 16 combinations are written in ordinary language as digits 0 - 9 and letters A - F.

Logical Operations

The basic operation involved in literature searching is the matching of search requirements (one or more index headings or descriptors) against the index headings or descriptors in the document. In manual or machine coordinate indexing systems the descriptors in the question and in the documents are translated into numbers or patterns of holes and a search consists of the matching of these numbers or patterns. Most searches require the matching of more than one set of numbers or patterns in a specified way. The type of matching which can be done has already been described as logical product (search for descriptor A and B), logical sum (search for descriptor A or B), and logical difference (search for descriptor A but not descriptor B). Another decision has to be made whether only the combination of descriptors is to be matched or whether the matching has to take into consideration the stated relationship among descriptors.

Form of Search Results

The results of a machine search can be the serial numbers of ducuments of probable interest (printed out or interpreted on punched cards), the bibliographic citations, abstracts, or even full copies of documents of probable interest, either printed out by the machine or reproduced photographically by the machine.

C. Organization of Section

Since the purpose of each of the installations is the same, namely to provide pertinent documents in answer to a stated or anticipated need, attempts have been made to include the same type of information about each installation and to arrange it in the same way. The planned arrangement of each description of an installation is outlined:

Machine characteristics: A brief description of the machine and any modification or change which has been made from the standard model of

the machine is included.

System environment: Information about the user, the collection, the indexers, and the starting date of the system is included.

Index: The type of index which is used, the size of the index vocabulary, the control which is exercised over the index vocabulary, the specificity of the index heading, and the depth of the index are mentioned.

Code: The type of machine code uses is indicated.

Use: The kind and frequency of use is reported.

Claims: Any cited advantages of this system over any other system, conventional or coordinate, are indicated.

Evaluation: Comments about the system, either reported in the literature or made by the writer of the present report, are given.

For most installations, information about one or more of these points was not reported. This is mentioned in many cases since the lack of information about an important point such as the use of the system will have a bearing on its evaluation.

Electronic Information Systems

Section D - Photoelectric Systems

Part 1. Systems Using IBM-Type Punched Cards and Sorted by Photoelectric Methods

Samain's Electronic Selector

Samain, in the 1940's, attempted to move away from the limitations placed on punched cards by fixed field coding and to develop a system using Hollerith cards which would provide greater versatility. This system has been described by Samain in several articles (9, 10) and by other writers in briefer descriptions (11, 12). Several foreign patents and one U.S. patent (13) have been granted to Samain.

The system provides that each of the twelve rows of the cards is divided into two portions, making 24 sections of 40 positions each. Each 40-position section can represent a six-letter code word. Each letter is coded by one or two punches in a six-position segment of the 40-position row. The extra spaces can be used to record logical or syntactical relationships (14). The coding system gives the card a capacity of twelve words of thirteen letters, or twenty-four words of six letters, or thirty-six words of four letters. An artificial six-letter vocabulary then has the capacity of 60 million different words (15).

A special typewriter is used to record the terms by making perforations in the cards (16).

A selector is provided which is:

> easily adjustable to any term and which reads the cards one by one at the speed of approximately 400 cards per minute. The reading will be performed with photo-electric cells by a special process. With successive selections, we can select the cards relating to any group of terms (17).

Shaw, in his review of various information systems, indicates that photo-electric cells were not used but that:

> In searching for information on the card, brushes were used by Samain to make contacts through each point at which the hole had been punched, just as in normal electrical Hollerith searching, but these pulses were fed into an electronic memory, which stored the pulses. When the pulses match the combinations of pulses set into the memory as the subject of the search, that particular card is dropped into a pocket, just as is the card in normal Hollerith searching (18).

He states that:

> This mechanism was an interesting experimental development, which has never been pursued to a conclusion as to whether it may usefully be applied, and the inventor has given it up in favor of a modification of the Rapid Selector (18). (See Filmorex)

IBM Photoelectric Scanner

The IBM photoelectric scanner, known as the "Luhn Scanner," or "Luhn Machine" has been described in detail by Luhn (19) and in a number of reports in Chemical and Engineering News (20, 21, 22). The system used, like that of Samain, does not require fixed field coding and operates by photoelectric processes (18). In this system, the code consists of five punches in the twelve positions of the IBM card column. The combinations derived from the five punches in the twelve punch

field:

> are grouped in a number of series which are assigned to sets of alphabetic, numeric and special characters, including lower and upper case, as available on standard typewriters. Certain series represent 2-digit numbers by a single combination (or pairs of letters) thereby cutting space requirements for such information in half as compared with present IBM card coding. Division marks for separating words are part of the letter code, an arrangement favoring compactness of recording (23).

In scanning, inquiry cards are punched with a complementary pattern of the seven holes not punched in the document cards. The cards are passed lengthwise past a photoelectric scanning station at the rate of 1,000 cards per minute. When the opaque portions of the inquiry card match the holes of the codes on the document card, a blackout results and the mechanism is activated to drop the card into a special pocket. One photocell scans from one to four columns and the photocells may be wired to act independently or to produce various logical combinations (24).

A switch on the card punch enables it to punch either the "question" or "answer" card. A sorter and transcriber have also been developed as accessories to the system. The machines are of the same order of complexity as standard IBM machines, and in their early development, it was expected that they would not cost any more than standard IBM machines (24).

The five-hole codes yielded 792 possible combinations per column, thus allowing for complete upper and lower case alphabets, two-digit numbers, special symbols and operations, and one hundred two-letter combinations (25). These two-letter combinations were worked out in a series named the "Luko" series (26).

The prototype equipment was completed in 1950 and first publicly demonstrated in New York at the World Chemical Conclave, September 1951. The ma-

chines were then experimented with by Perry at the Center for Scientific Aids to Learning at Massachusetts Institute of Technology where it was reported that they were highly satisfactory for searching purposes with the speed limited only by the maximum rate at which the cards could be handled, that is 600 per minute (27).

Methods were worked out whereby a group of cards could be scanned as a unit by means of a hold-over device. This allowed for the simultaneous scanning of a group of cards which pertained to a single document. Counters were also to be added to the machine to indicate the number of cards satisfying a certain criterion or a combination of criteria being searched, so that one pass of the cards could select a generic category and at the same time count the number of cards in each sub-category (21).

A patent was granted in 1955 to H. P. Luhn for the photoelectric device for scanning cards. The device described in the patent contained two scanners, each capable of covering four fixed-field locations on the card. The patent claim indicated that desired cards are selected by means of a relay device operated when the scanners receive no light rays through the combination of master card and specimen card (28).

The IBM photoelectric scanner was not put into production by IBM because of the cost of incorporating the five-hole code. The new equipment which would be required with five-hole punching, it was felt, would not be in enough demand to warrant the cost, even though the process worked satisfactorily. It was decided to revert to the standard punch sorter with three punches and regular IBM transcriber with standard IBM punch codes, but to continue using the Luhn scanner re-designed for greater flexibility in searching for complex relationships. It was reported that the new scanner was to be delivered to Perry's group at the Battelle Institute early in 1955 (29).

Shaw states that:

> This attempt to sort over the total area of the

> card was not very successful because of the mechanical problems in handling the card so as to control the passage of light, with close enough tolerance between the punched card and the inquiry card to make this effective. The machine also provided only about one-half of one per cent of the speed of the Rapid Selector. It is no longer in process of development (18).

The FID Manual also describes the system and states that a second model was to be built "which may prove to be the production prototype" (30). Shaw, in 1956, stated that the machine was an experimental model developed by Luhn "on which development work was discontinued more than a year ago" (31).

A report from the Welch Medical Library Indexing Project stated that they ran a trial run on the IBM Photoelectric Scanner in which 140 articles were indexed and coded.

> The results of the trial run were highly satisfactory. This machine seemed to have great possibilities and we would have liked to have had more experience with it (32).

Since machines were not available, all future work on this project was done on the IBM 101 (31).

The principles of the Luhn scanner have also been described by Taube in introducing the principles to be used in his "COMAC," or "Continuous Multiple Access Collator," described later (33).

The Bush Patent Office Report refers to the Luhn system as the IBM X-794 and states that it:

> is a special machine which utilizes a machine scanning code of 792 characters that is recorded on standard punched cards.... Questions can be combined, up to a limit of 72 characters, in various logical combinations such as any one or more, any three or more, and all

> of one or more except certain specified codes. Cards responding to the question are automatically selected and accumulated, and, where several questions have been combined, the accumulated cards are then separately segregated for each question. Operating speeds, for the simultaneous search of the entire card, are expected to be about 1,000 cards per minute (34).

Taube points out that even at 1,000 cards per minute, it would require 16 1/2 hours to search a million items to answer one two-termed question. He states that because of this:

> a searching system, even when so advanced a device as a Luhn scanner, can only be used for relatively small collections or for collections which permit the division of items into mutually exclusive classes, each one of which is small enough to make searching the total class practical (35).

He states that:

> The great advance of the Luhn Scanner was its demonstration that free field coding could be used with punched cards and that one card could constitute the question which interrogated the store on other cards (36).

Continuous Multiple Access Collator

Taube has recently proposed a new machine, which he called the COMAC, or Continuous Multiple Access Collator, based on the principles of the Luhn scanner but which will be efficient enough for use with a collection of a million items (37). The system to be used is like that of the Luhn scanner except that collation as in coordinate indexing is to be used instead of linear searching (38).

The COMAC operates under the principle of matching codes on one punched card, against codes on another

punched card and punching the code for the logical product code on a third card. This card can then be collated with a third subject card and the final answer printed rather than punched. Thus, there is no refiling necessary. A special code known as the "Chinese binary" which allows numbers from 1 to 999,999 to be punched in two columns of the standard IBM card is used. By using 74 columns, thirty-seven item numbers can be punched on a single subject card (39). With a collection of one million items and an average of 20 descriptor terms per item, assuming 10,000 terms are used, the 20 million items resulting could be punched on 540,540 cards. The 540,540 cards, in this system, would be organized into 10,000 groups, one for each subject, averaging 54 cards to a group. Using the uniterm system, a two-term question could be answered by comparing two 54-card groups. Cards could be added at any time without "dedicating space" for them in the group (unlike Minicards). It is stated that the 3 million patents in the Patent Office could be handled on approximately 1,600,000 cards, and again assuming a 10,000-word vocabulary, they would average 162 cards per group (40). The procedure in collating would be to advance the cards endwise two columns at a time (41).

It is stated:

> We have not attempted in this paper to describe the Comac apparatus. However, from our studies of existing punched card equipment, binary to decimal converters, comparators, etc., it appears that once the basic concept of the Comac is accepted, the construction of a device for single code comparison represents only a very modest development effort. Actually the character of the physical equipment necessary is practically deducible from the new concept of collation as a matching and print-out process of item codes rather than a card selection and interfiling process (42).

It is assumed that cards can be advanced two columns at a time and compared in the COMAC at about

twice the rate they can be advanced in existing card reproducers that feed cards the long way. With COMAC cards containing 37 codes, the 54 cards containing 2,000 codes could be read in 100 to 150 seconds. By doubling this figure to allow for intermittent advance of the two groups of cards, it is estimated that it would take three to five minutes to search the average two-termed question. This is compared with 16 1/2 hours for the Luhn scanner in a similar situation. It is expected that in any sizable installation there would be a number of COMAC machines available so that numerous searches could be carried on at the same time. Thus, if there were five COMACS available, they could answer one question per minute (43). *

The COMAC was conceived under a research contract with the Air Force Office of Scientific Research (44).

Addendum: The machine version of the COMAC, hereafter known as the IBM 9900

Special Index Analyzer (IBM 9900) is described in an IBM publication (45). The machine is composed of 3 units:

> A modified IBM 36 Card Punch which is used for reading cards;
> A logical and intermediate storage unit which contains both the control equipment, a paper tape punch, and a paper tape reader for retaining the intermediate results of operations;
> A typewriter for automatically printing the results of the search (46).

The basic storage unit of the system is an IBM

---

*There is no indication in this report of the type of hardware to be used or whether the collating is to be done by mechanical systems. Since it is described as a development of the Luhn machine, it is placed here with photoelectric devices rather than with strictly electronic searching systems.

card. Each card contains the following information: the encoded name of the descriptor, the sequence number of a descriptor card in a descriptor deck, and the encoded document serial numbers. Each document serial number is encoded on 6 columns of the card. A maximum of 12 document numbers can be encoded on a card; the remainder of the space is needed for descriptor card sequence number (47). (Taube, as we have seen, put up to 37 document serial numbers onto one descriptor card (48).)

New documents are added to the system by assigning them the next available serial number and adding this serial number onto the pertinent descriptor cards. One card is made for each descriptor which applies. The descriptor code and the document serial number are added to each punched card. An x punch is punched in column 1 to identify the card as a new document number card. The new document number cards are added to the regular descriptor deck by means of a logical sum operation in the IBM 9900 (49).

In an IBM 9900 search, 2 decks of descriptor cards are matched for common document serial numbers. The first descriptor deck has to be reproduced on paper tape before it can be combined with the second descriptor deck, which is on punched cards (50). Searches can be made for logical products, sums, and differences (51). Only 2 decks are matched at any one time. The results of searches--the matching document serial numbers--appear on paper tape. The numbers are then printed out on a form. The search time for a typical search is indicated. A logical product search with 3 descriptors which include 300, 240, and 360 documents respectively requires about 10 minutes machine time (52). If we assume that the average descriptor in this collection includes 300 documents and if we use an average of 20 descriptors per document and a descriptor vocabulary of 10,000 descriptors--Taube's figures (53)-- then we can calculate the approximate size of such a file.

$$\text{Average number of documents per descriptor} = \frac{\text{Total number of documents x average number of descriptors per document}}{\text{Total number of descriptors}}$$

or

$$\text{Total number of documents} = \frac{\text{Total number of descriptors x average number of documents per descriptor}}{\text{Average number of descriptors per document}}$$

$$\text{Total number of documents} = \frac{10,000 \text{ x } 300}{20} = 150,000 \text{ documents}$$

A 3 descriptor logical product search for a file of about 150,000 documents takes about 10 minutes of machine time, as compared to Taube's figure of 3 to 5 minutes for a 2-term search of a file of 1,000,000 documents (54).

Use: The IBM 9900 was demonstrated at the International Conference on Scientific Information in Washington, D. C., November, 1958. No reports of the use of the machine are as yet available.

Part 2. The Rapid Selector

Early Development

According to Shaw (55), the first practical application of electronics to selection of data on film was probably that of Dr. E. Goldberg of Germany as revealed in the U. S. patent granted 29 December 1931. Goldberg's patent was applied for on 5 April 1928 (56). It claims:

> A process of carrying out adding, sorting, statistical and like operations which consist in exploring indications upon a search element comprising a search plate and a record element comprising a record card or strip and causing the radiating energy to actuate a recorder when the explored indications upon the search plate and record element are identical, the indications of one of said elements being penetrable by the rays and the indications on the other element being impenetrable by the rays (56).

Dr. Vannevar Bush of the Massachusetts Institute of Technology is credited with developing the basic principles of organization of knowledge applied in the Rapid Selector and the basic electronic system involved (55, 57, 58, 59). An experimental machine known as the "Bush Rapid Selector" was worked on at the Massachusetts Institute of Technology and was announced there in 1940 (60, 61).

The Bush Rapid Selector:

> made selections of particular bits of data from a checkerboard of light and dark squares on each frame of film that formed a code. Each square had to have a photoelectric cell to mon-

> itor it. When a particular subject was wanted the machine was set for a certain pattern and when all cells in that pattern received light impulses, there was the desired item.
>
> The drawback to this system was the number of photoelectric cells required (57).

The machine was designed for abstracts of much greater brevity than used in the later Rapid Selector. "The file speed was low enough with respect to the rate of advance of the recopying camera to obviate the need for any slowdown on the film drive in the case of closely spaced 'hits'(62)." According to Shaw (63), the principle used "was not operable and the selector failed to work."

The machine was dismantled when in World War II it was necessary to salvage the electronic parts used in the machine (57).

> After the war a number of interested persons again took up the problem of the Rapid Selector. Among them were Ralph Shaw, Librarian of the Department of Agriculture, and a group of persons who had originally worked with Dr. Bush, and were now independently organized as Engineering Research Associates, with headquarters in St. Paul (57).

The need for finding more efficient methods of handling reports was of particular concern to the Office of Technical Services and the Department of Commerce. The problems were discussed with Dr. Vannevar Bush, and Ralph Shaw and the Engineering Research Associates proposed a development contract on the Rapid Selector. The Department of Commerce approved this contract under the Office of Technical Services in 1947 and $75,000 was allocated for the work. Engineering Research Associates were to produce the prototype machine under the direction of Ralph R. Shaw, who was also to prepare information for and to test the machine (10).

Operation

The fundamental principle of operation of the Rapid Selector is the application of a photocell to determine when a desired code, made up of white and black spots, matches codes identifying abstracts of articles filmed on 35 mm. motion picture film (65, 66). The Rapid Selector used 2,000-foot rolls of 35 mm. film containing approximately 100,000 pages of text or abstract, and space for 600,000 coded index entries (63).

> The information is stored on one-half of a standard 35 mm. frame and the binary coding representing its subjects is stored in the form of opaque dots photographically produced on the other half of the 35 mm. frame. Holes representing the subject looked for are punched into an opaque mask, and the film is run over this mask. Light passing through the film and the mask falls on a photocell and as long as light reaches the photocell nothing happens. When the black dots on the film match all the holes in the interrogator mask, no light reaches the photocell for a tiny fraction of a second, this permits the stroboscopic camera to operate and a projection print of the frame of text associated with the code dots is made while the film keeps moving along and continues its search operation. The stroboscopic camera developed for the rapid selector makes its picture in two-millionths of a second so that there is no need to stop or slow down the film. Thus searching is done at the rate of 120,000 choices per minute and copies of pertinent pages are made as the searching is done. The machine was designed to use available auxiliary equipment; the width of the image on the take-off film is 0.9 in. so that the film taken from the selector may either be read in a microfilm enlarger or run through an automatic 35 mm. V-mail type enlarger and converted into full-size paper copies at the rate of about 500 full-size enlargements in seven minutes (63).

When frames which are to be photographed are too close together, the mechanism which moves the next frame into position for exposure cannot move quickly enough to have it ready in time to photograph. This problem was solved using an anticipatory head or photoelectric scanner which anticipates the approach of a frame which is too close to be photographed at high speed. This device slows down the speed of operation from 300 feet per minute to 50 feet per minute to allow photography of the second "hit." After the second picture is taken, the machine resumes its normal speed (67, 68, 69).

Descriptions of the recording camera and auxiliaries, optical systems, electrical control, and mechanical design can be found in the Engineering Research Associates' "Report for the Microfilm Selector" (70), and in an article published in Electronics (71).

After the original Rapid Selector was developed, a high speed intermittent camera was developed, following a suggestion of Vannevar Bush. The mechanism was developed by Ralph Shaw and a group from the National Bureau of Standards headed by Jack Rabinow, through funds provided by the Atomic Energy Commission (72).

Film Preparation

In preparing the file:

> When the operator photographs the abstract he simultaneously enters, on a keyboard device, as many as six separate catalog descriptions by using the code numbers. These appear on checkerboard form on the film right beside the abstract. Photographing is a routine, rapid job, and can be done without professionally-trained help (68).
>
> The coding of an abstract is in the form of seven-digit numbers. The abstract originally entered on the microfilm can be described and classified by a maximum of six such seven-digit numbers. Any abstract may then be se-

> lected by interrogating the device with any one of six characterizing seven-digit code numbers as an index (73).

In photographing materials for the machine, the operator:

> has before him a book in which all the subject categories are listed. Beside each category a number has been placed. Because each reference has room for seven digits, up to ten million number combinations are possible. Rearrangement of film could, of course, give more possibilities. Therefore, several reference systems can be adopted "bodily" by the machine, and used side by side; the laborious task of recataloging so as to adopt a single set of subject headings to a collection is thus avoided (68).

Evaluations

The performance characteristics reported (74) for the Rapid Selector were as follows:

Speed, ft./min.:	300
Speed, abstracts/sec.:	180
Speed, 7-digit numbers/sec.:	1,100
Reel capacity, feet:	2,000
Reel capacity, abstracts:	72,000
Running time, min./reel:	6.7
Recopying camera, frames/sec.:	30
Minimum reel slowdown time, sec.:	0.12
Film passed after a slowdown command, inches:	4

The Patent Office Report, in comparing the Rapid Selector with other devices, stated that it could search in four to six minutes material which would take an ordinary card sorter 900 minutes and an IBM 101 type sorter about 72 minutes.

It was believed that the machine would be capable of 500 feet per minute operating speed if 1. larger mo-

tors were used driving the reels, or 2. the film length between the anticipatory projectors and the main projector were doubled (76). It was believed that the system could be used for filing and indexing letter files, literature of the type included in a publication such as Chemical Abstracts, and for improvement in the index in such services, and in searching of Patent Office files (76). The inventor stated that:

> while useful results may be achieved merely by using the machine to do more speedily and more efficiently what we can now do...a really important contribution to the advancement of science will result only if we can re-think the methods of organization of knowledge to take full advantage of the new technique (77).

Shaw believed that with high reduction ratios and inclusion of bibliographical material only, instead of abstracts, the Library of Congress catalog could be stored on film, representing only 2.5 cubic feet of storage space and about forty minutes of running time if the entire catalog needed to be run. He believed that if the reels were mounted in cartridges of five second runs, it should be possible for one machine to answer at least twelve questions per minute or 8,000 per day and that "one machine might then, if properly used, handle all the reference uses of the public catalog of a research library" (78).

Coding and Selector Design

While the Rapid Selector was under development, there was considerable discussion of coding for the machine. Wise and Perry (79) suggested adapting coding system devices for key sort cards to coding for the Rapid Selector. This system would employ code designations of six letters each instead of the seven digit ones used, thereby increasing the number of available code designations from 10 million to more than 300 million. It was proposed that these codes be superimposed on the Rapid Selector coding area, permitting as many as 16 contexts to be coded on a microfilm frame rather than the limitation of six. It was believed that this

would increase the ability to conduct searches for combinations of several concepts at one time instead of permitting only single-concept searches. It was believed that this coding system would require relatively minor changes in the design of the Rapid Selector.

Mooers (80) criticized this suggestion on the grounds that it was inefficient in its utilization of available code space, its high rate of false drops and its inflexibility in requiring a constant number of code marks for each concept. He believed that random coding would equalize code mark frequencies and would be far more satisfactory. Wise (81), in a reply to Mooers, defended multiple code word coding and claimed that word coding is flexible, does not require a code dictionary, can show relationships between ideas, and can be made random by utilizing symbols or a modified alphabet.

Proposed Modifications

It was estimated by Engineering Research Associates that the Selector could be duplicated for about $50, 000. It was also reported that the company was considering the development of a "junior model" applicable to collections of 5, 000 to 10, 000 items. This model would have simpler coding and scanning and would not photograph the microfilm selected but would stop the scanning operation so that the selected frames of the microfilm could be viewed by the operator (82).

According to the Patent Office Report, Rapid Selector improvements are being designed at Yale University, using the same principle of 35 mm. film with both code and text, but:

> the equipment is designed to be complementary to standard punched card equipment. The code area permits simultaneous search of up to 400 columns, or the equivalent of 5 punched cards. Recording of the code information is achieved through transformation of the pattern of holes on a card to a pattern of lighted miniature bulbs that photograph as black dots. The in-

> terrogation system uses a system of plug-in phototubes which provide greater latitude in the selection of coded data by searching any pattern involving any columns or by searching a range of codes within a column, if desired.
>
> The microfilm selector equipment, then, has definite advantages for searches where the desired result is the facsimile reproduction of text, drawings and other graphic material, or abstracts. Its chief disadvantages are that the medium requires a serial selection, that reproduction is a separate process of development, that the area for coded selection is limited by the space required for the text material, that the coded entries cannot be changed or added to without recopying or splicing the entire roll, and that the roll imposes a fixed physical grouping of the material which must be determined at the time of recording (75).

The same report states that:

> In the Yale version of the Rapid Selector, a magnetic recording tract strips of the film has been suggested to carry pulses produced by recognition signals from the Interrogator Unit in order to activate the recopy camera. These pulses would be retained on the film after the interrogation run. This technique would make possible the transfer of the film to a separate printer so that rearrangements of selected frames could be made up into separate reels as desired. It is certainly possible that the portion of film now reserved for optically sensed recording of coded index entries could be replaced by an area for magnetic recording of similar selection information that could be more readily revised (83).

In 1956 Shaw discussed the Rapid Selector and made proposals for its improvement. He pointed out that in spite of the fact that the Rapid Selector used imput rolls of 2,000 feet of film containing approximately 100 thousand pages of text and 600 thousand index

entries in a quarter of a cubic foot of space, the input unit required four minutes of running time plus another minute or two of reel change time so that only ten reels could be run per hour and that if even only one reel could be used per inquiry, only 48 questions could be answered per day despite the high speed operation of the electronic parts of the Selector. He pointed out that the lack of balance in input, internal running speed, and output of the machine invalidates the value of the speed of the electronic selector. He believed that, "the basic error was the assumption that we could run fast enough to avoid preclassification" (84). He suggested the use of preclassification, using 50-foot cartridges instead of 2,000-foot rolls, suggesting that the search time could then be changed from 6-minute units to half-minute or one-minute units. He stated that, "This requires additional development work but the principle has been established and there appear to be no special difficulties about this development" (85).

Shaw stated that modifications have been made in the Selector which have eliminated about two-thirds of the electronic equipment and half the optics required in the first model and that these have provided for complementary as well as direct coding. He stated that:

> the product obtained from the rapid selector was adequate for use even in its first stage, and the cost of the machine appears to be one that can be brought within the range of any of the research libraries that have need for any mechanical devices or systems. About seventy thousand dollars were spent on the first prototype; about fifteen thousand dollars additional were spent in simplifying the mechanism and eliminating unnecessary elements under new theories developed from operation of the machine. In its present form the machine could probably be duplicated for about $10,000 to $12,000, but enough additional knowledge has been accumulated in the course of experimentation with the machine in operation that it appears probable that another $40,000 to $50,000 invested in development work should

> yield a production model of the machine that could be duplicated for around $20,000, and thus would not only be much faster and would supply full cycle response to questions but would be cheaper than any of the other machines available or in process of development. The development of a new rapid selector should be based on high reduction ratio microfilming, at least at the 60-diameter level, so as to carry the full article in the single frame opposite each set of code dots. Probably slit photography should be substituted for flash photography because that would reduce the amount of electronic equipment required and would increase the reliability of the machine still further, while reducing the number of elements that might require maintenance work (85).

According to a recent report, work is currently being done on the Rapid Selector at the Bureau of Standards.

> As part of the investigation of various information retrieval methods and machines, the Division is attempting to evaluate the practicability of the Bush-type Rapid Selector. Some of the opto-mechanical parts of the third machine were obtained from Yale University as a basis for further work. To overcome some of the difficulties found in the earlier machines, certain recent advances in electronic computer technology and in information handling are being incorporated in the equipment. The laboratory has built a small interrogator and comparator for use with the modified Yale film transport and copying equipment. The system will use binary code words of 40 bits each and as many words are required, the limit being set by the amount of electronics desired. At the present speed of operation, the equivalent of 240 punched cards' worth of code information can be searched and copied per second. A test film has been made containing about 1,000

> articles from the NBS Technical News Bulletin. The machine master film has been copied on mylar base material which has far greater strength than the acetate base materials. Some test loops have passed through the machine many thousands of times at 60"/sec. without breaking. The individual parts of the system have been checked out; debugging the entire system for reliable operation is in process (86).

Shaw had pointed out early in the development of the Rapid Selector that, "while substantially unlimited in storage capacity, in speed of reproduction and in range of selectivity, (it) is limited as to the number of different transactions it can carry out in a given time from one set of instructions" (87).

Bedford, in reviewing machine systems for handling information, in 1956 stated that:

> Microfilm is high in storage capacity, but the convenience of the unit record and ease of manipulation is sacrificed. The retrieval process is not complete until the second film is developed and the enlarged print is secured. The Selector is undoubtedly an excellent mechanical system for fairly static historical collections, but it is not feasible for rapid access to a rapidly growing diverse collection subject to shifts in requirements (88).

Vickery, in commenting on the Rapid Selector, stressed that high-speed selection devices perform only a portion of the total process of information retrieval, that of scanning symbols. He believed that when the other factors are considered, the reductions of storage space and searching time achieved by the Rapid Selector are bound to be considerably less than is claimed, and that significant advantages of speed occur only for searches yielding hundreds of references. He opposes the idea that the great searching speed in the machine solves in any way the intellectual problems of indexing, classifying and coding (89).

Patents

The patents which have been granted for the Rapid Selector include, "System and Apparatus for Selective Photographing" (Ralph R. Shaw) (90) which covers various aspects of the Rapid Selector apparatus and operation; "Means for Eliminating Interference Between the Optical Trains of a Photographic Reproducing Apparatus" (Lawrence R. Steinhardt) (91) which describes a device which would prevent fogging of the copy film in the Rapid Selector by having the light coming through the code area of the master microfilm be of a different wave length than that shining through the text area, making possible better shielding of the copy film by use of light wave length filtering; and "Photographic Apparatus" (Ralph R. Shaw) (92) which describes the Rapid Selector camera used for photographing various page sizes onto the text portion of the microfilm while simultaneously photographing the code pattern of a fixed size onto the code portion of the microfilm.

A number of technical reports, in addition to the one quoted above (58), have been issued by Engineering Research Associates (93, 94, 95). Additional references on the Rapid Selector are listed in the bibliography (96-103). The Rapid Selector has also been described briefly in a large number of more general publications.

Part 3. Filmorex

The Filmorex System has been developed by Jacques Samain in Paris and has been described in a number of pamphlets (104, 105, 106) and articles (107-111). The system uses rectangular pieces of microfilm 72 x 45 mm. which are divided into two sections, one section containing the code which is searched using photoelectric cells and patterns to match the codes desired, and the second section containing the document or an abstract thereof.

The Film and its Preparation

The Filmorex card is a heavy weight acetate sheet coated with photographic emulsion, 45 x 72 mm. (112). (The current card size is described as 70 x 45 mm. in at least one reference (113), and 60 x 35 mm. in two references (114, 115)). The Filmorex cards are prepared from a continuous roll of film thirty meters long (113). A double lens camera with a special keyboard is used to photograph the two sections of the film. After development, the film is cut into the card size. If additional copies are wanted, they can be reproduced from the film prior to cutting (116).

Coding

The coding area is twenty columns with each column capable of recording a five-digit number. Each digit is recorded as a pattern of two opaque dots and three transparent squares (117). In another description, the coding area of the card is divided into 25 parallel columns, each divided into six groups of six positions, making it possible to record a six-figure number on each line and 25 six-figure numbers in the entire coding zone (118).

Selection

Selection is done by passing the cards through a photoelectric scanning device which examines each line of the coded area consecutively. Whenever the desired codes are encountered, the scanning device operates an apparatus which sends the card into a special pocket. Those cards which do not possess the desired codes are sent into a second pocket (119). The selector reads 600 cards per minute (4, 000 digits per second) (120). One reference indicated 700 cards per minute (119), and Shaw states that the running time is 400 cards per minute, theoretical speed (112). Samain states that difficulties lie in the smallness of the code squares, which are approximately one square millimeter, and the speed of passage, which is approximately 1, 000 impulses per second, and because of the irregularities of the film. Reliability is achieved by having five reading stations each with a photocell, an amplifier and a thyratron controlling a relay (121). The system will work with the microfilm cards in random order (119) or the selector itself can be used to sort the cards in any desired order by successive passes. Duplication of cards into classified groups will speed the selection process by making it necessary to pass fewer cards through the selector (122).

Use of Filmorex Cards

The document or abstract can be read directly with a standard microfilm reader or it can be reproduced or enlarged by standard processing methods (119). Since the cards are not reproduced automatically, the original cards have to be taken from the file for use so that it cannot be used for searching until the cards are refiled (112).

Use of the System

It is stated that:

> As of the present time, all biological papers abstracted in the "Bulletin Analytique" have been encoded (123).

In the same paper it is stated that the Bulletin Signalétique, which publishes about 130,000 abstracts a year in the fields of mathematics, physics, chemistry, and biology, is using the system. It is stated that:

> Instead of preparing an index, like Chemical Abstracts, we decided to initiate in 1954 a bibliographic research device which provides information with regard to questions posed by searchers (124).

Coblans states that Samain has been experimenting since 1954 at the Centre National de la Recherche Scientifique (CNRS), Paris, with the Filmorex system.

> For the purposes of the CNRS each entry in the Bulletin Signalétique could be transferred to a micro-sheet with coding of all of its subject aspects, author, periodical, etc., and the reference number of the abstract in the bulletin. In this way a list of all reference numbers of abstracts in the bulletin on a requested subject for a specified period of time could be supplied as part of its service to subscribers at a nominal charge (125).

Coblans states that the mechanical aspects of recording and selecting can be done but that the major problem is classification, "the elaboration of a system viable over the whole gamut of knowledge and enabling selection of all abstracts relating to a definite subject" (125).

Filmorex Claims

It is claimed that the advantages of the Filmorex system are 1. the cards are very strong and can pass through the selector many thousands of times, 2. the cards are small but hold much information, 3. the selection process is simple and almost fully automatic, 4. the selection speed is very high, and 5. the cost of the equipment is relatively low (126). It is stated that a decimal classification system is used but that the coding system is not rigid (123). It is stated that the selector can search for various logical combinations of

five ideas at one time (121).

*

Part 4. Minicards

The Minicard system for documentary records and control was first described at a meeting of the American Documentation Institute in November 1954, and in published form in January 1955 (127). A later report described the equipment used in detail including photographs, tables of operating rates and flow charts of operation, and described the filing and searching operations of a hypothetical collection of one million Minicards (128). The developers of the system believe that it will have broad application and that it will be useful for computer and business applications as well as for handling all types of documentary information (129).

The Minicard

Minicards are small pieces of photographic film, 16 x 32 mm. in size. Near one end of the piece of film, a slot is provided which makes it possible to handle the Minicards on metal sticks (130).

The space of the Minicard is divided up between code areas and image areas. The image areas, of which there can be a maximum of twelve, each record the equivalent of a legal size page. When twelve image areas are used, there is some space for coded information but as more space is used for coding, less is available for text (130). The reproduction in the image areas of the Minicard is done at a reduction ratio of 60 to 1. Thus, the Minicard has a much higher storage capacity than microfilm (131). In the earlier report, the maximum digital information capacity of the Minicard when it carries no graphic images is given as 70 columns of 42 bits each, or, 2940 bits in all(130). In the later report, the code capacity is specified as 66 columns of 43 binary digits each, which are used to make up 7 six-bit characters plus one "parity checking"

bit. One of the 7 characters in each column may be used as a "tag" to signal the kind of code of the other 6 characters in the column. A boundary signal may be entered in the open field of the card to indicate linkages or relationships between codes (132). Duplicate Minicards are made for each significant code entry in order to reduce searching time (133).

Minicards are handled on sticks which are used in presenting the Minicards to the machine and for entering and removing the cards from the files. The sticks have a capacity of 2,000 Minicards (130). The sticks of Minicards are combined into file units consisting of either 10 or 100 magazine units (134).

It is claimed that Minicards have roughly the same cost, card per card, as punched cards, but on the basis of digits or bits per card, "Minicards have a cost advantage over punched cards which amounts to a factor of many times" (135).

Preparation of the Minicard

In making Minicards, the camera performs two basic functions: 1. the exposure of code patterns and 2. the exposure of document images. The film in the camera is in roll form which is cut into separate Minicards after processing. The code, which is in the form of alpha-numeric characters, is punched onto paper tape. The paper tape is used to control the camera which then exposes the code pattern automatically. It is also possible to enter the code directly from a keyboard to the camera or to enter it from punched cards or magnetic tape. After the codes have been exposed, the document pages are exposed to the film (136). The camera can record forty to fifty 6-page documents and their codes per hour (137). This operation requires an operator who positions the documents. Otherwise the operation is largely automatic. Both line and continuous tone copy can be reproduced. The film is processed in roll form and then passed through a "film chopper" which cuts the film into individual Minicards, and stacks them on the Minicard stick. In duplicating Minicards, a contact printing procedure is used

with a roll of film. Additional coding can be entered during the duplication if desired (136). The negative Minicards are used to produce positive cards with code designations as required. The negative cards are put into a master file, which may be arranged by accession number, and the positive cards go into the working file used for selection of material (138).

Minicard Sorting and Selection

The Minicard system contains equipment for sorting the Minicards and distributing them to their correct file locations. For normal sorting, the cards to be filed:

> are fed one at a time past a reading station which sorts on one digit at a time in a designated code column, the cards being directed to one of ten receiving magazines. Sorting can be done on both numeric and alphabetic characters (139).

A "fine sorter" has two sets of ten receiving magazines, each with a reading station preceding it, arranged in a closed circle. The Minicards are transported around the circle so that it is possible "to program this machine to sort successively to any number of digits without attention by an operator" (140). In a "locked sorter" the reading stations and magazines function in the same manner but are moved by a linear transport mechanism. With this sorter, file magazines can be attached to it and sorting is done directly into the file. In the earlier report it is stated that sorting and selecting can be done at the rate of 1,800 cards per minute (130). The later report lists the scanning rate for the filing sorter as 1,000 cards per minute and the scanning rate for selection as 1,200 cards per minute (137).

In the selection of Minicards, questions are coded and Flexowriter tapes are prepared for use on the selector with limits of 20 words per question. Plugboards are used for logical relations and conditions (141). It is also possible to set up boundary specifications for

recognition of "less than" or "greater than" (134, 142). The Minicards from the storage magazine pass a reading head where the data is read from the Minicard code field (143). In the selector, the Minicard code field passes a fixed array of 43 photocell detectors (134). The data is examined in the electronic circuit of the selector and, when the data is recognized as satisfying specifications of the question asked in the selector, the Minicard is directed into a separate receiving magazine. Minicards not selected pass into a different magazine (134). Output of the system is in the form of duplicate Minicards or full-size prints of the selected master Minicards, the selected cards themselves never leaving the system but being returned, after reproduction, to the working file (144).

Thus the Minicards will be used either by means of viewers of various types, from which the text on the Minicard can be read directly, or through enlargements provided by automatic enlarging devices (145). The user in either case keeps the copies which have been reproduced for him (144). The duplicator produces Minicards at the rate of 120 per minute and the enlarger processor makes 300 prints per hour (137).

Minicard System Claims

It is claimed that the Minicard system has the following properties important for documentation:

1. The system handles graphic and digital information in one record medium. The Minicard, a discrete record unit, has a high information capacity and a high activity capability.
2. The system has an efficient record duplicating capability.
3. Document information in the system may be delivered directly.
4. The system provides the input-output convenience and file space advantages required for large files.
5. The system has capabilities for organizing files by machine. This makes possible relatively

short searching time for large files.

6. The system has search capability to satisfy the requirements of present information systems (146).

Minicard Development

On 30 June 1955, a demonstration of "prototype Minicard equipment" was held for consultants and representatives of government agencies in Rochester, New York. At that time it was indicated that the equipment was scheduled for delivery for use by the Air Force early in 1956.

> The various items of equipment demonstrated included:
>
> 1. A step-type camera for recording previously prepared codes and documents on a continuous film.
> 2. A processor to develop exposed film.
> 3. A film cutter, for preparing Minicards from the continuous film after development.
> 4. A Minicard sorter for simple sorting of Minicards. This unit performs operations similar to those of conventional punched-card sorting equipment.
> 5. A Minicard scanner, or a prototype machine was used to illustrate some of the operations that this unit will be able to perform once design and construction work have been completed (147).

Lewis and Offenhauser in 1956 stated that:

> Many of the details of the Minicard system are still under study; this is true especially of the specialized sorting and classifying equipment associated with its coding aspects. The photographic aspects... appear to be substantially complete (148).

In 1957 Eastman Kodak representatives stated that the Minicard system had not yet been tested in full-scale operation, that it is still in a development phase

and that no decisions have been made as to making Minicard equipment available commercially (149).

A report by Hawken states that:

> Minicard has made considerable progress in the last year. One complete unit has been installed in Washington and is undergoing testing. Two other units have been delivered to other government agencies. My informant, Dr. Feldman, stated the first unit would be used in intelligence work (probably Pentagon). He was not at liberty to disclose the destination of the other two units. A fourth unit to be completed this summer will be retained by Eastman at Rochester and will be available for experimental work to organizations interested in the potential use of Minicard for special applications of their own.
>
> The extent to which the physical, chemical, optical, sensitometric, electronic, and mechanical problems of Minicard operation have been solved is most impressive. More and more work is being done to make the system more flexible and capable of handling more sophisticated problems. But Dr. Feldman admitted that the main problems in the effective use of the Minicard system are those of classification common to all systems, on which coding and ultimate retrieval depend (150).

Cost of the System

Published accounts by the developers of the system do not indicate costs of equipment or operation. Shaw, in 1956, stated that:

> Something on the order of one and one-half million dollars has been spent or allocated to this development program and it is not yet operational.
>
> It has been determined that a second machine can be built during the process of developing the first one for $350,000 and that if the

> machines were produced in batches of 100 complete sets, the price of the production run would be of the order of $150,000 per installation (151).

Evaluations

Shaw points out that:

> This system is based on the theory that the Minicards selected will be reproduced by contact printing and then will be restored to their old stick, so again the material is not available for immediate researching (151).

Taube, under a contract for the Office of Naval Research, has made what he calls "A Case Study in Document Storage and Retrieval" of the Minicard system. He admits that his analysis could not be checked against actual test results since the system had not yet reached a full operational stage. His analysis indicates that the high reading speed and compact storage of the Minicard system are greatly weakened by the systems of coding and filing employed. The selector rate of 1,800 Minicards per minute would require over nine hours for a sequential search of a file of one million items, thus necessitating duplication and prefiling if the operation is to be economical. In prefiling, he points out that at least one file magazine (a stick with a 2,000 Minicard capacity) must be dedicated to each term or class in the system. Otherwise, duplicates in the output will result from filing Minicards for more than one class on one stick. The unequal loading of the ministicks in having unused space for lightly used terms and requiring extra magazines for heavily used terms, thereby creates problems for interfiling and searching. The hundred magazine file units require only one cubic foot of storage space, but their weight of fifty to one hundred pounds which must be transported to and from the selector prevents their being stacked, thereby drastically reducing the claimed space saving. His calculations indicate that the Minicard files and machines will occupy about as much space as the original file. He points out another shortcoming in the placing of both

text and code on the Minicard, which requires the output of one search to be refiled before the next search can begin and which inactivates the selector mechanism while the text is passing under the reading heads. He believes that the extensive capacities of the selector may be more elaborate and thereby more expensive than necessary to achieve a satisfactory output (152). Taube and Heilprin have also issued a report relating the size of questions to the work accomplished by a storage and retrieval system. Using the Rapid Selector and Minicards as examples, a study was made of sequential and parallel searching and the effect of multiple simultaneous searching on continuous and discontinuous systems (153).

Shaw states that:

> There appears little in this development in terms of its potential application to storing material in research libraries, locating it and reproducing it. There have, however, been several developments in the course of this project that are of considerable significance, including the development of 60-diameter reduction equipment for filming and reproducing from film. The balance of the operation is not much better than that of punched cards. Again the machine and make-ready costs are so great that this system will be of value only with highly repetitive routine processes, which, at least so far, cannot be anticipated in research libraries. It is a possible substitute for Hollerith type cards in installations that now use them, since it will operate somewhat faster than Hollerith cards and will save a great deal of space as compared with punched cards. It is, in effect, a punched card system using smaller cards and the Rapid Selector's sorting system (151).

Bedford has evaluated the Minicard in her review of systems of information and storage retrieval (154, 155). She states that:

> The Minicard combines the high storage capacity of microfilm with the desirable characteristics of the punched card for ease of manipulation, prearrangement and sorting (156).

She states that in terms of the criteria developed for this review of systems:

> The Minicard system can be rated as an excellent method for the organization of information. The system is infinitely expansible in that it is completely open-ended and it can hold information to the depths required for all kinds of information. Any changes required to meet new and unforeseen developments can be made quickly and easily (156).

In summary, she selects the Eastman Kodak Minicard along with the International Business Machines electronic data-processing system as the two systems which have emerged from an over-view of machine literature searching in the United States as feasible for installation in the project with which she was associated (157).

In another project carried on by Kent, Perry and Egan, a study has been made of Minicard indexing. These reports apparently deal only with indexing and not with development of the Minicard system (158).

Part 5. Miscellaneous Photoelectric Systems

Photographic Glass Disks

The 1954 Patent Office report states that:

> Research and development on photographically recorded disks storing digital information in binary code form is in progress at Eastman Kodak, International Telemeter Corporation, Logistics Research, Incorporated, and other laboratories (159).

According to this report:

> The use of photographic glass disks is estimated to provide storage densities of the order of 100,000 to 1,000,000 bits per square inch. A disk that is 6 inches in diameter might store 10,000,000 bits, or the information equivalent of 150 feet of 35 mm. film. On a band 4 inches wide at the outer edge of a glass disk 16 inches in diameter 20,000,000 bits might be stored, all of which would be accessible to a single reading station. Using a combination of flying spot scanning techniques and rotation of the disk itself at about 800 r.p.m., it should be possible to read this digital information at data rates of up to 1,000,000 bits per second, or the information equivalent of about 2,000 conventional punched cards in one second. This rate would be approximately 3 or 4 times faster than the rates for magnetic tape so far reported. However, magnetic tapes can be written on, and hence revised, at the same data rates as for reading, whereas the writing process for the photographic media are much slower and involves the separate process of expo-

> sure, developing and fixing.
>
> The probable access rates to photographic disks are very promising, since the entire disk might be read in about 5 seconds as against the 5 to 6 minutes necessary to read through the conventional computer tape or the Rapid Selector microfilm roll. In the work at International Telemeter, the combination of index entries with text material has also been studied using the same principle of integral index that is used for the Rapid Selector (160).

Photographic techniques for information storage have been studied by International Telemeter Corporation (161, 162). They reported that storage devices with a total capacity in the range of 10^7 to 10^9 bits, incorporating favorable random access features, were possible and that the use of color photography could multiply the data storage capacity by three times. The photographic medium would begin with punched cards. The fastest system considered would employ a flying-spot scanner searching a moving storage surface such as a rotating disk or drum. The flying-spot scanner would also be used with the photographic storage. It is believed that the combination of a small eraseable storage with a large photographic store will make available a device which will make it possible to apply modern computing and information processing techniques to practical non-numerical problems (161).

Recall Film Index Systems

The Patent Office Report also refers to a film index system which is closely related to minicards. This system, developed by Recall Incorporated, would use Kalfax film. This film is sensitive to ultra-violet light and has a heat developed emulsion which can be processed in normal light. The system provides film cards which record reduced images of the document and coded index entries which are produced by contact printing from a punched card (163).

Film Library Instantaneous Presentation

The Benson-Lehner Corporation has announced an automatic microfilm searching machine whose purpose is to search for a particular frame on 16 mm. film at speeds of 300 to 600 frames per second and then to present this frame to the operator viewing. A binary code, in the form of black bars on a clear background, is used, the code utilizing 32 bits of information. The film reels used are 1,200 feet long, containing 72,000 frames. The code capacity is eight descriptors. Location of the documents is by accession numbers which must be in numerical order on the film reel (164).

Automatic Micro-Film Information System

The Automatic Micro-Film Information System, known as AMFIS, has been under development by E. A. Avakian for some time. A mimeographed report dated 2 September 1952 has been abstracted as follows:

> AMFIS is the first realistic attempt to materialize Bush's dream of Memex. AMFIS employs a mechanism that can store either 1 1/2 million documents or 9 million catalogue cards. By keying document accession number on a ten digit key-board operator can project legible image onto a viewing screen at center of desk, or at remote points, within two seconds. Documents are reproduced in less than half a minute by standard techniques such as photostat or xerography. Reports may be microfilmed on 8 mm. and/or 16 mm. film. 20-inch strips of microfilm are inserted into a holder, which is scroll-like in arrangement and holds 1,000 such strips, which may be replaced in less than 30 seconds. The desk viewer is standard and the remote viewer employs a flying-spot scanner. The basic mechanical components were patented under the title of the Stored Function Calculator, U.S. Patent No. 2,610,791. The unit described uses 16 scrolls but more could be added. Browsing is possible and similar in operation to standard microfilm readers. AMFIS undoubtedly is a unique approach to the unidimensional character of film.

> Its possibilities for reference and catalog departments of large libraries should be explored, as well as in information storage problems where quick retrieval is essential (165).

The patent referred to describes the equipment used for storing the microfilm slips and for projecting particular frames to a viewer upon request (166).

Another report on AMFIS (167) is referred to by Lewis and Offenhauser (168).

In a recent report, AMFIS is described as a device for combining great reduction of storage space with great speed of access. It is indicated that the intent is not to have an information retrieval machine but a system that will select a given document whose identity and call numbers are known. By dialing the call number of a document, the required film is moved into position so that it can be read on a screen or a full size copy made. It is suggested that if a television scanner were substituted for the light, the document could be transmitted over long distances. It is stated that any size of film can be used and that the machine can also be adapted for work with micro-opaque material. It is stated that the capacity of one AMFIS machine is several million microfilm frames which may be stored either singly or in strips, allowing for the addition or deletion of frames as desired. This report states that the system's "final realization merely awaits concrete financial assistance" (169).

Magnavox Film Data Recorder

Magnavox has developed a film recording system known as the Magnavox Film Data Recorder for storage of digital and pictorial data together. "It involves a reel of microfilm either 16 mm. or 35 mm. with coding space in selected frames. The coding space covers one-half of a single 16 mm. frame or about one-fourth of a 35 mm. frame. A total of 90 bits of information are stored in each coding space.... A machine has been developed which can search, select, and display specified frames as determined from the coded area

information. The selected frames can be enlarged and copies reproduced both of the image data and of the coded data" (170).

E. Punched Card Sorters

1. IBM Electronic Statistical Machine, Type 101 (IBM 101)

One of the first uses of the IBM 101 multiple column sorter was for the Welch Medical Library indexing project. In their final report on machine methods for information searching (171) the authors indicate that IBM suggested the use of the IBM 101 in the fall of 1951. The machine was delivered early in 1952 and was used for several experiments at that time. Mention was made in this report of the use of the IBM 101 by the U. S. Patent Office and two publications by members of the U. S. Patent Office staff on machine literature searching, both presented in 1951 (172, 173), make references to a code used for a machine-based index but do not discuss the IBM 101.

In a survey published by Kent in 1958 (174) of non-conventional retrieval systems, 9 IBM 101 installations are included. These are:

Battelle Memorial Institute
Ciba
DuPont*
Merck, Sharp and Dohme (Westpoint, Pa.)*
Proctor and Gamble
Schering
Smith, Kline and French*
Socony Mobil*
Union Carbide Chemical*

Installations marked with an asterisk are described in separate publications and will be discussed below.

Characteristics of the Machine

The IBM 101 can recognize and sort for one or more punched positions in one or more columns in one sort through the machine. The rate of sorting is 450 cards per minute. The machine is programmed by wiring a plug board. This plug board can be supplemented by a dial board to facilitate programming selected cards and can be directed into any one of 12 pockets, the 13th pocket being for rejected cards.

The full machine, i.e. the machine with all internal wires connected, can sort up to 60 of the 960 positions of the card in one sort. Sorting can be programmed for a logical product, a logical sum, a logical difference, or any combination of these logical operations.

Information to be sorted on the IBM 101 can be entered on the punched cards by direct code, by indirect code (numeric or random superimposed) and by free field code with some modification of the machine. The card passes under the reading brushes of the IBM 101 while moving in a direction parallel to the vertical columns. The presence or absence of any hole or combination of holes in each of these 80 columns can be determined only after all 12 horizontal rows of the card have passed the reading brushes and effected the transferred or non-transferred status of a series of recode selectors or relays. After the last row of punches in the card is sensed and before the next card is fed the relays are electrically tested for conformance with a preselected pattern of holes. If agreement is found the card is deflected into 1 of 12 pockets rather than into the reject pocket.

The IBM 101 systems described up to the present time have merely listed descriptors; relationships among descriptors in particular documents are not indicated.

It is possible to conduct more than one search at one time by directing search results from different searches to different pockets. It is also possible to conduct one search and to direct cards which only partially satisfy search criteria into different sorting pock-

ets. For example, a card which matches one descriptor is sorted into pocket 1, a card which matches 2 descriptors is sorted into pocket 2, and so on.

The machine has no internal memory. It can perform calculations on data entered on the punched cards. A printing unit is incorporated into it.

E. I. DuPont de Nemours Company
Chemical Department

This machine-based indexing system for internal research reports had been in operation for 3 years in 1957. A pre-installation assessment of the needs of the laboratory research personnel--the primary users of the system--indicated that an index had to answer a myriad of detailed questions in the chemical field and in bordering sciences such as biology and engineering. Ability to answer generic questions was also required, especially as applied to aspects of structure and elementary composition of chemical compounds(175). Responsibility for the selection of subject matter to be indexed is shared by the author of a report and the indexer, but final decisions are made by the indexer.

The index is divided into 4 parts: organic, polymer, inorganic, and miscellaneous (natural products, products of unknown composition, mixtures, "trade name" products, and information which cannot be tied to a compound or polymer). The following sections of the list of indexing terms apply to all 4 parts of the index:

Status of material being indexed (product, reactant, etc.)
Types of reaction
Reaction conditions
Properties
End use objectives
Modification or treatment (spinning, coagulation, coloring)
Miscellaneous headings

The index headings for structural features (functional

groups, ring systems, type of polymer, elements) vary for each of the 4 parts of the index, although some headings will appear in more than one part.

A list of captions (headings) and the corresponding codes is kept on punched cards in a "supervisable" (visible) file (176).

Code: Both direct and random superimposed codes are used. Direct codes are used in 56 columns for the most frequent headings (including document number). Random superimposed codes (4 positions per unit of information) are used in the remaining 24 columns of the card.

In addition to the detailed index to the reports, one card covering the major aspects of the subject matter is made out for each report. (It is not indicated whether this is a punched card or an index card).

Each compound and each polymer component is indexed on a separate punched card. This means that not all the index entries for a sought report will be on one card. A two-step operation is thus necessary, the first step being the selection of cards with a particular code, the second step being the matching of the cards for common serial numbers. If there are too many answer cards for visual comparison of report numbers, they may be put in numeric sequence by machine so that cards belonging to the same report may be grouped together (177).

Use: Frequency of use was not indicated.

Claims or comments: It is believed that the framework is adaptable to most types of information occurring in the reports being indexed. Any part of the index can be expanded to accommodate the additional needs. This index does not provide a unique identification of every compound since such completeness would demand a much more extensive system than is required for the size of installations contemplated. Usage thus far has given gratifying results since very few unwanted cards have been selected (178).

Merck, Sharp and Dohme

A machine-based index to published articles has been employed in the library here since 1950 (179). A Remington Rand sorter with a multiple (12 column) sorting attachment was first used. However, an IBM sorter is now being used. Approximately 30,000 references are included in the file at present. The library service is for a group whose interests touch on nearly every phase of the biological, chemical, and medical sciences (180). A staff of 2 professionals and 3 typists handles the input and output of the system.

The dictionary of indexing terms for the system consists of 3 parts:

General subject (an alphabetical list of about 1,000 terms)
Chemicals (slightly less than 1,000 compounds)
Diseases (terms from the American Medical Association's Standard Nomenclature of Diseases)

The dictionary of general subject terms contains cross references and instructions to the indexer for indexing under more than one level of heading, e.g.

Children	Also coded: humans
Humans	Also coded: children, when pertinent (181)

A 4-position random superimposed code is used for the more uncommon terms; a direct code is used for the common terms.

A dial board is attached to the sorter to facilitate programming the machine. For making special types of searches, a wiring system has been developed which will separate the 15 possible combinations which can be obtained from 4 descriptors into the 12 sorting pockets of the machine. Thus if the 4 descriptors are identified by A, B, C, and D, ABCD would be sorted into pocket 12, BCD into pocket 11, ACD into pocket 10, ABC into pocket 1, ABD into pocket 2...A or B or C

or D into pocket 9. Cards with a single position punched will have to be re-sorted a second time (182).

Mark sensing cards are used for the preparation of sets of index cards (183). False drops resulting from an interaction of the random superimposed code are not considered a problem. It is the goal to keep machine time for an "ordinary" search to 15 minutes. To do this, cards are divided by year and subdivided by types of papers, namely human clinical papers, veterinary clinical papers, experimental papers in the field of biology, and all other (184).

In an earlier paper (185) Mrs. Schultz points out some of the advantages and disadvantages of a punched card system. The advantages cited are flexibility in various respects; cards do not have to be filed in any order, answers can be obtained to specific questions without manual selections, and questions can be asked by any number of combinations of terms. The cited disadvantages are operational:

> The ability to search the card file depends on the proper operation of the machine. When the machine breaks down the system is temporarily inoperable.... Some searches can be made more rapidly with standard indexing. To find all the papers of a particular author would mean searching the entire punched card file (186).

Use: No data is given on the use of the file.

Smith, Kline and French

The system includes over 10,000 published and unpublished documents on a single drug, chlorpromazine, at Smith, Kline and French. Documents contain clinical and pharmacological information about the drug. The users of the system are scientists at the company laboratories (187). Highly trained scientists are not needed to operate the system, although the indexers must have scientific training to read the reports intelligently and to translate the author's words into code

words (188). In 1957, two years' experience with the system had been completed (189).

A separate code is used for clinical and pharmacological information but all data on a document is entered on one card (190). The pharmacological code consists of less than 260 index headings which are entered on 12 columns of the IBM card. The headings are categorized; direct (5 columns) and free field (7 columns) codes are used. A portion of the pharmacology headings with an indication of the type of code used may be given here:

<u>Subject</u> (direct code)
not specified, several
rodents
birds, amphibia
dog and cat

<u>Site of action</u> (free field code)
Central nervous system
 cerebrum
 hypothalamus
 brain stem
 spinal cord

<u>Special feature</u> (direct code)
unusual dose
unusual route
absorption

<u>Type of study</u> (direct code)
chemical
biassay
pharmacy

<u>Body system</u> (direct code)
metabolism of reference drug
blood, hemopoietic
cardiovascular

<u>Type of action</u> (direct code)
reference drug on function
reference drug on metabolism
reference drug on histology

The free field code makes use of 6 punches per column. This allows a total of 924 combinations per column, any one of which can be entered in one of the 7 columns reserved for this code. Any one of the combinations can

be entered in any one of the 7 columns. Since this code is sorted by pattern of holes on the card rather than by the location of the hole.

No detail is given about the clinical code except that the code is set up with a high proportion of direct punches or assigned positions so that data can be correlated and tabulated (191). The 6 position free code field can be sorted by wiring only one of the machine's selectors (192).

The results of the search are on cards on which the corresponding document number is punched and interpreted. The cards are either examined visually or used to print a list of IBM 101-prepared document numbers. References in the chlorpromazine file are also in Flexowriter tapes. Bibliographies can be prepared from these tapes (193).

Use: No data is given on the use of the file.

Claims: It is claimed that "There are no false drops, though in our experience the false drop is more a theoretical menace than an actual one" (194). The authors point out that this type of index is not limited to an IBM 101. It could be adapted to a card index, they state (195). Some of the other advantages claimed for the system over other systems are:

> It requires no decision by the indexer about the relative importance of the various factors.
>
> The subject heading list can be enlarged or made more detailed without need to disturb indexing already done or to upset the entire classification system.
>
> The code was planned to be applicable to any area of pharmacology or physiology (196).

Socony Mobil

The system at Socony Mobil includes about 95,000 articles and patents on petroleum technology indexed

since January, 1952 (197). A staff of a technical personnel selects, codes and classifies 12,000 articles and 4,000 patents per year. The system is primarily for the 330 technical employees of the research and development laboratory, although occasional calls arise from other divisions (198).

The machine installation actually serves three purposes:

1. Bibliographic cards (author, title, bibliographic citation and classification), which are prepared for each new reference on IBM cards. Weekly classified title lists are prepared directly from these IBM cards.

2. Continuous searches, which are prepared on requested subjects; bibliography cards are prepared weekly and forwarded to the 23 men who receive this service (199).

3. Retrospective searches.

Subject matter is encoded directly onto 41 columns of the card. The code is reproduced in a machine literature processing demonstration manual (200). Individual indexing units, the descriptors, are classified by broad categories such as unit operations, equipment, physical properties, and hydrocarbon chemicals. Since 2 or more descriptors are combined to form a search entry the descriptors are relatively generic. An example of descriptors used for a search is given by Crandall and Stumpf (201). The descriptors are:

Performance testing
Improvement by additives
Oxidation
Petroleum products
Lubricating oils

Since descriptors are only listed and their relationships are not indicated, some commonly used descriptors are repeated in two categories to reduce the

incidence of false combinations. For example, chemical elements are repeated under inorganic chemicals and under elements analyzed for.

The average machine running time for a complete literature search ranges from 5 to 6 hours according to the complexity of the request (202).

No overall statistics are given on false drops. The illustrated search (203) contained 24% pertinent references. What percentage of the non-pertinent references were noise and what percentage were false drops was not indicated.

In addition to the detailed index which is machine searched, the system is backed up with several auxiliary files, called satellite files, which are hand searched. This includes a trade name file and a principal code (subject) file. The latter file is used for spot searching where only a few articles are necessary.

Use: The system was used to conduct 33 machine searches in 1954, 41 machine searches in 1955, and 62 machine searches in 1956. An average of 45 manual searches per year was also conducted (204).

Evaluation: This is one of the earliest applications of machine literature searching in an industrial organization. The system has now been in operation for 7 years. Early experiments on a machine sorted punched card system were started as early as 1948 (205). It is a dual purpose system in that the cards are used for the weekly accession lists as well as for retrospective searching. The cost of the installation is consequently charged to both current dissemination of information and retrospective searching. Nevertheless the total cost seems to be high in terms of professional manpower, and the use of the information retrieval part of the system seems on the low side.

The use of one card per reference gives greater flexibility in sorting. That is, more descriptors can be coordinated in one sort through the machine and no subsequent sorting is necessary to identify matching

document numbers. But the use of one card per reference does result in a larger percentage of false drops because of the interaction of non-related descriptors. In view of the relatively low use of the system, it is probably more efficient to separate these false drops manually as is done in this installation rather than to try a more complex code in order to avoid them.

Union Carbide Chemical Company

Internal reports and patents are included in the machine literature searching installation. The number of documents in Union Carbide's system is not indicated in the 1957 reference but the file is said to consist of over 20,000 cards and to contain some half million entries from patents and internal reports (206).

Both direct and random superimposed coding are used to encode subject matter. Specific subjects, namely individual chemical compounds, are identified by a 5-digit serial number followed by a single digit to indicate the role of that particular chemical in the document. These roles are:

Reactant
Product
Catalyst
Chemical agent
Materials of construction
Physical agent
Negation (removal or absence of chemical)

The presence in the document of analytical methods or physical properties is also indicated by means of such role indicators.

The 6 digits for the specific subject (5 digits) and its role (1 digit) are entered directly into a 6 column field. Index entries for other than specific chemicals are entered as 4 positions in a 10 column random superimposed code field. These entries are called concepts and are exemplified by terms such as aldehydes, olefins, and oxidation (207). About 1,500 of these concepts are used.

The punched card is divided into 16 code fields, 7 of which are used to encode subject matter. Six 6-column fields are used for encoding up to 6 chemicals and their roles per card. The 10 column field is used for the random superimposed coding of concepts. Since a specific subject can be entered in any one of the six code fields without any prior order, the matching appears to be by code pattern rather than by position.

Use: No data is given on the use of the installation.

Claims:

> Every single piece of information in this file can be literally read in less than an hour as often as requested.... The cost of a single retrieval in technical personnel time is less than one tenth the cost of a conventional search regardless of the number of subjects covered (208).

2. IBM Electronic Statistical Machine Type 101 (IBM 101) With Row by Row Scanning Attachment

An IBM 101 with a row by row scanning attachment has been described by Luhn (209). This was demonstrated at the International Conference on Scientific Information at Washington, D. C., in November, 1958. The basic coding unit sorted with this machine is the 80 column row; there are 12 such rows per IBM card. Each row can accommodate 12 six-letter words in Hollerith code or even more in a more condensed code. Indications of relationships among descriptors can also be brought out in the code.

The machine is programmed by wiring a plug board. Searches are made for patterns within a single row of a card. One row is searched after another. The incidence of a match is stored until all 12 rows on a card have been scanned. Searches can be made for the presence of any one of any combination of several of the desired patterns on the same card (210).

Results of the search (document serial numbers) are printed out as the search progresses so that the order of the cards in a file need not be disturbed (211). This is a requirement if 2 or more cards are made out for any one document, since relationships among cards would otherwise be destroyed.

An experimental system making use of this machine at the IBM Research Center is briefly mentioned by Luhn. The index was prepared entirely by an IBM 704, which was fed a machine-readable transcript of the documents. (The computer was programmed to select indexable information and to translate it into index language.) A "nodal index" was used in the experiment. This is an index which includes for each key word (des-

criptor) all of the other key words which are found to have been paired with it in the original texts (212).

Use: No information is given on the use of the system.

Evaluation: The IBM 101 with row by row scanning attachment has two advantages over the conventional IBM 101:

1. Relationships among descriptors can be incorporated into the code.

2. The code is no longer restricted to the fixed capacity of a conventional column by column sorted punched card.

Row by row sorting allows punched cards to be handled as a continuous unit, as does magnetic tape. Against these advantages must be set the following disadvantages stated in the report:

> Each set of cards, comprising a single record, must include two extra cards, one containing control instructions for the machine, and another to act as a spacer card (213).

Thus if there is an average of 2 cards per document in the conventional IBM 101 system and 4 cards in the IBM row by row scanning attachment system, then the 2 extra cards per document would require in effect double the sorting time.

3. The Universal Card Scanner

The Universal Card Scanner is another machine which was demonstrated at the International Conference on Scientific Information at Washington, D. C., in November 1958. This machine and 2 literature searching applications in IBM libraries are described by Luhn (214). These appear to be the only literature searching applications, since the machine is not commercially available at present.

The operation of the machine is described by Luhn (215):

> The Universal Card Scanner (UCS) scans cards fed through it in a manner similar to that employed on conventional punched card sorters. It is capable of discovering whether any or several of a given set of patterns are wholly or partly contained in any of the record cards scanned. This function is performed by a no-pulse matching process under the control of a question card which contains prototypes of the patterns sought, likewise represented by punched holes. This is the adaptation of an electronic method to the optical principle of 'matching by blackout,' employed in an earlier IBM card scanning machine, frequently referred to as the 'Luhn Scanner'. As was the case in the earlier mode, the present machine features the use of a punched IBM card (Question Card) for furnishing the patterns to be searched for in a record file.
>
> The particular matching process employed in the UCS requires that the pattern on the record cards be given in complementary form, i. e.,. the various marks or elements of the pattern need to be represented by the absence

> of holes and all else by the presence of holes.

The entire card is scanned in one pass through the machine.

> The machine scans a card as a unit, i.e., whatever is contained within the twelve positions of the card columns is treated as one continuous pattern and a match or lack of match is determined once per card on the basis of 12 such position patterns. Patterns may be of any width desired and a plurality of them may be recorded across the card (216) at predetermined locations, either adjoining or overlapping each other.

Several types of codes are mentioned for possible use with the machine: the standard alpha-numeric Hollerith code (217), random superimposed code (218), and word coding (words used in their original form), made more efficient by spreading the words over a larger portion of the card using the first letter of the word to indicate the starting column (219). Randomized square coding is also mentioned. For this code successive letter pairs are used. These letter pairs are marked as the intersections of rows and columns, where a particular row stands for the first letter of a pair and a particular column for the second letter. A method called chain spelling is used here; it consists of linking the pairs by repeating the second letter of a pair as the first letter of a succeeding pair. This chain is closed on itself by an additional pair by end-around spelling of the last letter and first letter of the code word. The code word TUG, for example, is spelled TU, UG, and GT. Relationships among words, i.e., the coding of 2 words as a phrase, can be indicated by a modification of this code. If the code words TUG and DEV are to be connected they would be encoded as TU, UG, GD, DE, EV, VT. Individual words can still be identified by disregarding end-around letter pairs (220).

Searches can be made for logical sums, products, and differences (221). A search can be programmed to

sort cards with different degrees of code matching into different pockets. For example, cards which scored no code match are sorted into the reject pocket, cards which scored one code match are sorted into pocket 1, cards which scored 2 code matches are sorted into pocket 2, and so on (222).

Two applications of the UCS are described. One experimental system is used for technical reports in the technical library of the IBM Military Products Division at Owego, N. Y. The system supplements a title and author card catalog of technical reports. Descriptors are entered on IBM cards as 3 letter random squared codes, as discussed above, in any one of three 12 column fields. (Only one of the 3 fields has been used so far; the other 2 fields are devoted to bibliographic information) (223).

The code is entered on the record card (document card) by machine duplication of a dictionary card. The dictionary card contains the dictionary term in word form. Housekeeping information and the particular pattern of punches are repeated in three 12 column fields. Machine duplication can be made selectively in any one of the 3 locations on the record card.

The question card is divided into six 12 column fields. Code patterns are entered into the first 3 fields by machine duplicating corresponding dictionary cards. The code patterns in the 3 additional fields are also duplicated by machine but the dictionary card is turned around (the former right edge pointing to the left) and thus a mirror image is produced in the field. This mirror image is compensated for by appropriate wiring of the control panel (224).

The second application at the Information Retrieval Research Department of the IBM Research Center deals with literature on information retrieval and machine translation. The installation is similar to the first example with one major exception. The encoding operations were carried out entirely by an IBM 704. A card or cards are made for each author, title, and source of document. Descriptors are selected by the

machine from titles of documents. By means of a table look-up, a pre-determined set of insignificant or "common" words is excluded from the titles. The remaining words are considered to be significant and useful as descriptors. These words are listed for each document and are stored on magnetic tape. Record cards are prepared by the IBM 704 for each document (225).

Use: No information is given on the use of the systems.

Evaluation: The UCS may be compared with the IBM 101 since the machines are similar in size, cost, and capabilities. The UCS has several advantages over the IBM 101. Its sorting speed is said to be 1,000 cards per minute (226), which is over 2 times the speed of the IBM 101. Randomized square coding, a pattern rather than a position code which is not readily applicable to the IBM 101, makes very efficient use of card space. The space thus gained on the card can be used for entering bibliographic information. Randomized square coding also permits the indication of relationships among code words. Programming of the machine by inserting a punched card is faster than the control panel wiring required of IBM 101 systems. But additional control panel wiring is also mentioned (227) and thus the actual saving in time for programming might be small. Both the IBM 101 and the UCS are capable of handling logical products, sums, and different types of searches. The ability of separating cards by the number of matched codes can reduce the number of times a deck has to be sorted. This can be done with both machines, but in the case of the IBM 101 this requires rather intricate wiring.

A drawback of the UCS is that it is a single-purpose machine. It does not have the flexibility of the IBM 101 which can also be used in statistical work and can do a certain amount of printing.

4. Interrelated Logic Accumulating Scanner (ILAS)

In 1951 an experiment with one of the first row by row scanned punched card searching systems was presented at an American Chemical Society meeting by members of the U. S. Patent Office staff. (Samain had suggested the use of row by row scanning instead of conventional column by column scanning to obtain greater flexibility as early as 1945) (228). The paper given by the Patent Office staff members--it was later published (229)--is interesting for several reasons: a mechanism is provided for linking and sorting for related descriptors in the code; use is made of specific descriptors with more generic descriptors in the form of a classification system built in. This last technique was subsequently suggested by Perry (230) and called preparation of abstraction ladders. It forms the basis for the ILAS.

The operation of the original row by row scanner is described by Andrews in the Patent Office Research and Development Report #6:

> The machine tested each row of holes for a desired pattern before the next succeeding row of holes reached the reading brushes. If such tests were affirmative the transfer of a relay took place and its transferred state continued for all succeeding holes of the card as well as for an indefinite number of succeeding cards until such time as the relay ultimately was released by a specially located control punch in one of the cards which would then cause the sorting of the last card (into a particular pocket other than the reject pocket). By this means' twelve rather than one logical decisions were made for each card but what is more im-

portant is that coded data of any length could be processed as a single record without being confined to a single card (231).

In June 1956 a Census Bureau 488 Multi-Column Sorter, a machine which has many features in common with the commercial IBM 101, was delivered to the Patent Office. This is the machine which was completely rebuilt into what is now known as the ILAS. The machine consists of the card sorting unit and a connecting console in which the logical circuitry is housed.

The ILAS contains 80 column relays which detect the presence or absence of a hole under each of the card reading brushes 12 times during each card cycle. Each column relay has 12 independent sets of contacts having normal (no hole punched) or transferred (hole punched) positions so that 12 different combinations of 80 punches may be detected by proper interconnection of all corresponding sets of contacts on each of the 80 relays. Making these connections by plugboard would be a very difficult job, and one almost impossible to check. Instead, rotary switches are used. These rotary switches are mounted on the front panel of the console. A hexadecimal code is used for descriptor codes programmed by the rotary switches. The hexadecimal code allows 16 combinations (descriptors) from 4 positions, which in this case are 4 positions in a 4 column field. Each rotary switch has 20 positions on each of the 4 levels and makes the interconnection between common, normal and transfer contacts of corresponding contact sets of 4 successive column relays. This means that each of the 16 positions of the 16 descriptors in the 4 column hexadecimal code field can be set with the rotary switch. In addition, one position serves as a shunt for all 4 column relays with which it was associated and another position serves as a shunt for the one 4 column field which it controls as well as all the remaining column relays of the row which is under the control of the switches (232). These last two positions facilitate programming. Only rotary switches which are to transfer will have to be changed (provided that the others are in the shunting position); and one rotary switch can be used to control all succeeding ro-

tary switches in a row if a change in the program requires the deletion of all codes in a particular row.

The punched card used for this experiment is divided into several parts:

1. Signal:

> The first four columns of each horizontal row provide a 'Signal' field in which is punched any one of 12 different combinations of holes to be used as markers. For example if each row of holes in the card is considered to be a 'word' consisting of a number of different hexadecimal characters, then a group of such words separated by a distinctive signal could be a 'phrase,' a number of phrases could be grouped to make a 'sentence,' and several sentences grouped to form a 'paragraph.'

This is quite similar to what Perry calls barriers (233).

> A distinctive signal would also be used to mark the end of all the codes representing each patent or document. These signals serve very much the same logical function that parentheses and brackets serve in mathematic expression (234).

These signals are programmed with a plugboard.

2. Modulant: a single hexadecimal character from the 4 position field in column 5 -- 8 is used to modify the remainder of the code word in that particular row. An example of modulant use is given in another Research and Development Report of the Patent Office (235). The code word "diazotization" is qualified by means of modulants to diazotizing agent, diazotizes, diazotizable, is diazotized, and diazotized compound respectively. A modulant field is provided for each of the 12 code words entered on the 12 rows of the card. A hexadecimal code controlled by a rotary switch is used to program for the modulant.

3. Subject matter: the code words (unmodulated and unlinked descriptors) are entered in columns 9--63. This area is divided into fifteen 4 column fields, each field being used for a 4 bit hexadecimal code. (A bit is a given position in the row.) Consequently, a code word of up to 15 hexadecimal characters can be entered on a single row. This area of the card is also programmed by means of rotary switches described above.

4. Interfix: the signal code groups code words into phrases, sentences, and paragraphs. This is useful, but still another grouping device is necessary. Andrews explains it:

> Many components must be represented logically as being in two or more different groups. A code representing such a component could not be physically located in more than one group (provided by the signal code) and repeating the code for each (signal) group would lose the relationship of the several codes as pertaining to but a single component (236).

To provide for a second order of relationships another coding device is included. This is the interfix. An example of its use is given by Andrews:

> A gear identified as 'A' drives shaft 'B' in turn drives pulley 'C'. If we code for A any one of a set of marks, such as '3' and add the same number to the code for B meaning that the driver-driven relationship is present if the two codes contain the same, although unspecified, interfix number. Similarly, B would be interfixed to C but a different arbitrary interfix number would be selected, such as '5'. The interfixed codes would then look like this:
>
> A--3
> B--3, 5
> C--5 (237)

In addition to the interfix code, A, B, and C would have

a common signal code to group the code words into a phrase. Interfixes are entered into columns 69 to 80 of the card. This area may be handled as a single field or as multiple fields of 2, 4, or 6 columns each. The interfix is programmed with a plugboard.

5. Hit relays: since coded data for any document does not have to be confined to a single card, some device has to be used to "remember" that part of the search specifications which has been satisfied. It must also instruct the sorter to alter the path of a card when all the search specifications have been met. In the ILAS, 24 hit relays are used for this purpose. Each hit relay has separate pickup and dropout connections and common normal and transfer contacts. There are also 48 interfix hit relays which are interconnected to "remember" each punch in the 12 interfix columns for up to 4 different code words interfixed in the same column (238).

Searches for logical products, sums, differences, and combinations of these logical operations can be performed on the ILAS. The speed of searching is 450 cards per minute, or the same speed of the IBM 101. On one occasion this speed was increased to 500 cards per minute before timing difficulties resulted in erroneous sorting (239).

The card deck has to be kept in exact order; otherwise relationships established among cards by a particular sequence are destroyed. Since all the positions on the card are used for the code and no space is provided for housekeeping information--identification of the indexed document--it is particularly critical that the order of the cards be not disturbed. In the modified IBM 101, the last card in a sequence is diverted in a sort pocket if it meets search specifications. This necessitates manual card interfiling of these cards before another search can be started. A print-out of the pertinent document number would have been the ideal solution but the cost of a printing unit was prohibitively high. Instead, a procedure called progressive sorting is used (240). Cards are fed into the reject pocket until the first hit is made. At this time the cards are

deflected into the next pocket until the second hit is made. The cards are then deflected into the next pocket, and so on until all the pockets are used. The bottom card of each sort pocket identifies the document sorted by the machine. The ordered state of the card file is maintained by restacking the cards from each pocket in sequence.

Use: Three applications of the machine are described. In a 1957 publication on the ILAS (241) Andrews states that the coding system described (with signals, modulants, and interfixes) is still in the formative stage but that the ILAS was tested with the deck of punched cards prepared on medicinal compositions by the U. S. Patent Office in 1950. Descriptors were used for ingredients (a physical admixture of 2 or more compounds), complex natural products, and functions (disclosures of uses, properties, physiological behavior). The ingredients were coded as specific descriptors with more generic descriptors built in across the row of the card. Each ingredient was assigned a number of codes indicating various characteristics of the ingredient such as structural groups, function, and source of natural product. The multiple codes for each ingredient were tied together by a grouping signal. The set of ingredients in a composition was tied together by another signal indicating end of composition. In 4. 5 minutes, 441 patents containing 6, 262 disclosures characterized by a total number of 18, 650 descriptive terms were scanned. This is a rate of 95 patents per minute (242).

The second application of the ILAS, described in another Patent Office Research and Development Report (243), is in the polymer field. It deals with ehylenic unsaturated homo and copolymers, classified in class 260, subclasses 80--94. 9 of the Patent Office Manual of Classification. The descriptor list includes the following types of disclosures:

1. Monomers used in polymerization
2. Inerts in the reaction
3. Solvents in the reaction
4. Catalysts
5. Processes

6. Conditions of the processes
7. Properties of the product
8. Uses

The disclosures are divided into 2 major subdivisions:

1. Ingredients--terms related to chemical compounds
2. Functions--terms including non-structure terminology, identifying processes, properties, conditions of reaction, and so on.

Each of the chemical compounds is given an identifying serial number. This number is entered on the card with a hexadecimal code. By using 4 hexadecimal codes in a field, a total of 65, 536 unique combinations can be obtained. To put it in another way, when 16 positions are used in a row any one number from 1--65, 536 can be obtained. In addition to this identifying number the genus--species relationship is given for each chemical. For example, in the case of the specific compound of ethylene the generic term mono olefin hydrocarbon would also be coded.

Functions of the chemical, e. g. solvent, catalyst, comonomer, are indicated in the modulant field by means of 2 hexadecimal characters. This provides a total of 256 unique modulants.

Related codes are linked together with a grouping signal. Thus, a phrase can be made out of the codes for ethylene, mono olefin hydrocarbon, and comonomer. Another grouping signal is used to connect all the codes which pertain to the same document.

Interfixes are used to show interrelations among various groups, for example the group for specific compounds and a specific process. Interfixes can also be used to show sequence in a process.

Numerical information is also coded by ranges. This is useful for indicating operating variables in a process. These variables might be pressure, temperature, or time.

Only half of the card, i. e. a 40 bit word per row instead of an 80 bit word per row, is used for this system. The first 4 bits (positions in each row) are used for signals, the next 8 bits are for the modulant, the next 16 bits are for subject matter, and the next 12 bits are for the interfix. The system is in operation but no data is given on frequency of use.

The third application of the ILAS is described by Leibowitz, Frome, and Andrews (244). The subject matter covered in this experiment is in the thiazine art. This refers to a group of sulfur-containing heterocyclic compounds. The entire 80 positions of each row on an IBM card are used to encode subject matter and relationships. The arrangement of the card is similar to that described above (245). Three levels of descriptor groupings are used:

> Ring systems or chains (a chain is a continuity of chain units consisting mostly of functional groups in an acyclic arrangement of elements).
>
> Compounds
>
> Patents

Each of these levels is connected with a grouping signal. Groupings within smaller units, e. g. ring to chain unit, are made by interfixes. The system permits variability in scope because:

> 1. The 'building blocks' of the system are small units. These units are separately and independently described, which permits the asking of a large number and a large variety of search questions for retrieval.
>
> 2. The descriptors are variable in scope--one descriptor may merely indicate a 6-membered ring while another indicates a positional relationship of heterocyclic elements or substituent groups....

> 3. Each collection of codes is gathered into a substructural entity. This permits search questions with respect to chemical compounds which vary in scope as desired with respect to selected portions of the molecule.
>
> 4. The groupings and interfixes provide the ability to specify relationships among the substructures and to obtain as much specificity as desired with respect to the compound search (246).

An interesting approach is the encoding of the compounds from the formulas by clerks. The system is being tested (247).

Evaluation: The Patent Office is experimenting with a number of systems. The ILAS is only one of them and as such does not represent the ultimate approach to the Patent Office's problem. There are some very interesting features in the ILAS. Most electronic systems for information retrieval formulated so far have been limited to a mere listing of descriptors. One thus gave up something very real when one departed from the traditional indexing system. In many cases the indexing was not so detailed, nor the subject matter complex, nor the size of the collection so large as to cause real trouble. The problem at the Patent Office demands a system which will permit precise specification of relationships among descriptors. This is made possible with the ILAS by means of the grouping signal and the interfix. There is no IBM keypunch now available which punches hexadecimal codes in rows of punched cards. A two step operation is thus necessary. This is somewhat awkward and increases the clerical cost of the operation. The sorting rate of 450 cards per minute is not very fast for a large file, particularly since 2 or more cards are used per patent (an average of 5 cards per patent in the first experiment) (248).

Andrews compares the ILAS with a computer:

> At the rate of 450 cards per minute 90 rows of holes or 7200 bits of information are scanned

> each second. This is approximately one twelfth of the scanning rate of commonly used magnetic tape input units (90,000 bits per second) but the fact that 12 sets of comparisons are made simultaneously on the fly without additional computational time further reduces the gap between punched card equipment and electronic computers costing at least 100 times as much as the ILAS machine. It should also be noted that the time required to prepare and load the machine with the data defining the search specifications usually takes only a matter of minutes and is a direct operation subject to visual checking whereas preparation of data for a computer usually must pass through several hands and, if trouble develops from the data, requires elaborate checking procedures.

He does add, however:

> The magnitude of the entire Patent Office searching operation is such that ultimately the largest and fastest available computer or specialized machine will be necessary (249).

F. Computers

1. Paper Tape

Western Reserve University Searching Selector

A pilot project sponsored by the American Society of Metals is now being conducted at Western Reserve University to demonstrate the feasibility and advantages of applying computers to the retrieval and correlation of metallurgical literature. The project consists of indexing 25,000 metallurgical publications over a 5 year period (1955-60) and testing the indexed file. Informative abstracts from Metallurgical Abstracts, the Journal of the Iron and Steel Institute, and Chemical Abstracts are used as sources for indexable information (250).

The indexing technique for this project was developed by Perry, Kent, and Berry and is described in their book on machine literature searching (251). Indexable information in a document is translated into an artificially defined group of symbols called a telegraphic abstract. Among these symbols are semantic factors, defined as

> ...a carefully defined set of generic concepts ...to serve both as reference points in designating the scope of information requirements and as a basis for designating important aspects of subject contents of graphic records (252).

A few hundred semantic factors sufficed as a basis for encoding a broad range of scientific and technical terms (253). Examples of encoded semantic factors are:

B CT	bacteria
B SR	absorb
M CH	machine
T TR	water

Semantic factors are supplemented with 1-letter codes to indicate permanent relationships; these are called analytic relationships, and are exemplified by:

A	class inclusion
E	material of composition
Y	attributive
X	absence of

Semantic factors coupled with analytic relationships are illustrated by the following:

MACH	thermometer is a member of the class machine
TXTR	anhydrous or absence of water (254)

Empirical relationships, called synthetic relationships, are denoted by a 3-letter code such as:

KAJ	starting material
KEJ	material processed
KOV	properties given for

In addition to the semantic factor and the 2 types of relationship indicators, the following devices are used in the telegraphic abstract: brackets, parentheses, and braces between symbols that constitute various designated runs such as phrases, sentences, and paragraphs. There are also other symbols to distinguish between codes for terms that are so closely related that they would have the same semantic factors or to distinguish a given individual whose class has been encoded. That is, they distinguish between different models of a machine or between a pyrometer and a thermometer (255).

The operation of the system is described by Kent, Melton, and Flagg (256). Published abstracts which are to be incorporated into the system are translated into

telegraphic abstracts. Abstracts are first translated into a standard form by eliminating the variations and complexities of the English sentence structure. This task is facilitated by making out a subject matter analysis form. The end result of this step is a telegraphic abstract on the worksheets. The next step is to code the individual terms and phrases of the telegraphic abstract with the aid of a code dictionary. The encoded telegraphic abstract is then translated into machine form by recording it both on punched cards and 8 channel punched paper tape. The telegraphic abstracts on 8 channel tape constitute the "library" which is scanned by the machine in a search.

The hardware used in searching consists of a modified Flexowriter and a panel of circuits, cable-connected with the Flexowriter. The code of the question is spelled out by appropriate wiring. The logic of the question is programmed by additional levels of wiring from one panel to another. Logical sums, products, differences, and combinations of these logical operations can be used in searching.

The operation of the selector is described by Kent, Melton, and Flagg:

> The Flexowriter 'reads' 8-channel punched paper tape in its reading unit. The punched tape contains the encoded library which is being searched. The panel of circuits receives impulses from the Flexowriter, according to which symbol punched on the tape is being 'read'. Therefore, if the letter 'A' is being read by the Flexowriter, the set of hubs on the Selector which correspond to the letter 'A' will receive voltage. If the letter 'A' appears in our question, we would have plugged a wire into an 'A' hub, so that when the voltage appeared, the logical apparatus of the Selector is put into motion. If the letter 'A' is not in our question, then voltage at the 'A' hubs will disappear or be 'washed out' as soon as the next symbol punched in the tape library is read.

> After the question on (for example) titanium has been programmed into the Selector the 'Send' button on the Flexowriter is depressed (which conditions the Selector to receive impulses from the Flexowriter) and then the button marked 'start read' is depressed. The search of the punched paper tape is thus initiated (257).

The result of the search can be a serial number of an abstract or paper, the bibliographic citation, or the entire abstract. The Selector--presently as experimental model to try out the system of encoding developed by Perry and his colleagues--is slow. Only 1 abstract is searched per minute. At one time, 10 questions can be searched (258). Procedures are being worked out in cooperation with the Eastman Kodak Company to encode the telegraphic abstracts onto Minicards (259). The use of magnetic instead of paper tape is also being considered in order to increase the speed of searching.

Evaluation: The WRU Searching Selector is thus far an experimental model which is admittedly too slow for searching. The system is intended for more sophisticated machines which, according to Shaw, do not exist. In a review of Perry, Kent, and Berry's book on machine literature searching Shaw states:

> The basic assumption that underlies this series of studies is that we have machines capable of literature searching.... As a matter of fact there are no machines in production that will even do a fraction of what is claimed here (260).

The telegraphic abstract-based system is also questioned because of the resulting large record of the index to a document in machine form. Shaw comments:

> It is difficult for this reviewer to see how much space is saved by using MUSRMACHTWMP 03 for the word thermometer. It requires 40 spaces to write Springfield, Illinois and 44 to

> write Chicago, so this notation would overflow from a punched card with only Springfield and Chicago needed on one card. The usefulness of a notation this long on a medium capable of storing a total of 80 characters only is open to question (261).

Punched paper tape instead of punched cards is now used by Perry as the storage medium; nevertheless Shaw's argument is still highly pertinent. The system is cumbersome.

The amount of detail which is encoded in illustrated telegraphic abstracts (262) is also open to question. The professional and clerical input cost of the system seems to be unduly high and in poor balance to the output cost. The large resulting machinable record appears to require a machine with an unnecessarily high searching rate, especially since we are dealing with an end-to-end search type of system. The detail of encoding appears to be based on the expectancy of searches for very detailed information. This type of search might be more efficiently conducted by restricting the machine search to the elimination of the bulk of the nonpertinent material and finishing the search manually.

Bendix G--15--D Computer

This computer consists of a Flexowriter--the input, output, and control unit--and a console which contains the arithmetic unit and the 2, 176 word (63, 004 digit) magnetic drum memory. An auxiliary magnetic tape memory can also be incorporated into the computer, as can punched card and magnetic tape input and output. The input and output operations proceed without interrupting computation (263).

Only one application of this computer to literature searching has been reported thus far. The work was done jointly by the U. S. Patent Office and the E. I. DuPont de Nemours Company. It is reported in ORD Report #13 (264), in which a code for ethylene polymer art to be used with either the ILAS or the Bendix G--

15--D is discussed. The basic code has been reported above under the ILAS section. The principal refinement introduced for the Bendix G--15--D is called the weighting procedure.

> Each subject group in a question set in assigned a relative numerical value, called weight, in accordance with its considered importance. During the search the machine operates in the following manner. When the subject is found for its first appearance in the disclosure its weight assigned to the subject is recorded. The weight is not recorded for any additional finding of the same subject. As each additional subject is found its weight is added to the previously recorded weights of other subjects. The total weight at the end of the search is compared with a minimum weight which has been assigned concomitantly with the questions. Assume, for example, subject A is assigned a weight of 1, B a weight of 2, and C a weight of 3. When the machine finds subject A, it records weight 1 and when it finds C it adds 3 to 1 and if it finds B adds 2 to give a total weight of 6. A complete answer to the question is A plus B plus C. If, in addition, answer B plus C is also acceptable, the minimum weight assigned would be 5 (265).

The results of the machine search are printed out by the computer. The print out for each document includes the relative weight of the document (maximum or minimum).

Use: A flow chart and program were written for the Bendix G--15--D Computer and an operable program was set up. Preliminary tests with a computer program have been encouraging (266).

National Bureau of Standards,
Automatic Microimage File

The Automatic Microimage File is a device for the rapid location and copying in enlarged form of microfilm images. The device is described in two very similar articles (267, 268). Images of documents or other forms of information are reduced to 0.1 inch frames of microfilm. On a 10-inch square sheet of microfilm, 10,000 of these frames are recorded. This sheet is termed a matrix. The locations of these frames on the 10-inch square microfilm are recorded as 20 bit numbers on a perforated teletype paper tape:

> The instrument is essentially a combination of digital computer electronic circuitry and a pair of precision servomechanisms that search X and Y axes of the matrix. The location of the desired frame is fed into a 20 bit (binary digit) register from the teletype tape. The register consists of a capacitor memory and coincidence identification circuitry. The first 10 bits recorded in the register control the Y position selection while the second 10 bits control the X position.
>
> The matrix is supported on a drum 10 in. in diameter and is fastened at one edge with dowel pins to insure its accurate location on the drum. The drum is servo-controlled in both linear and rotary axes of motion, corresponding to the X and Y axes of the matrix. The servos that shift the matrix to the chosen coordinates are mechanically coupled with precision gearing to two code commutators.
>
> The code commutators, one associated with each axis, control the coordinate positions to which the matrix is located. These commutators are photoetched with one hundred 10-bit numbers corresponding to the standard teletype binary bit code. The two particular positions on the commutators are selected by a serial mechanical search with contacting brushes until a code combination is found that matches the binary bits recorded in the 20-bit register. Magnetic clutches and brakes provide rapid starting and stopping of the drum with uniform overtravel in the location of every position on

> the matrix. A single induction motor supplies all motive power to the machine.
>
> At the beginning of the cycle of operation, a teletype tape reader reads a 4-decimal-digit number into the 20-bit register in terms of a binary digit code. A space symbol is customarily inserted in the teletype tape following each 4-digit number. On detecting this space symbol, the machine's program control stops the tape reader, engages the magnetic clutches on the X and Y servos, and looks for the compatible code on the two coordinate axes. When the compatible code is found, the clutches disengage and magnetic brakes stop the drum. A print lamp is briefly turned on to make a photographic exposure of the selected microfilm frame on the photosensitive paper. When the exposure is completed, the teletype tape advances to the next instruction, the drum returns to its zero position, and the machine proceeds to the next search cycle (269).

The Automatic Microimage File appears useful for storing documents in highly reduced form and for obtaining somewhat enlarged copies of these reduced documents. The images thus produced will have to be further enlarged before they can be read by the naked eye. The instrument is not a subject index searching device since it can only make copies of documents whose location is specified.

The Automatic Microimage File might have application in a 2-step searching system. The first step in such a system would be the subject search of the index. Document serial numbers would be the result of this search. The serial numbers would also indicate the location of the documents in the Automatic Microimage File. The second step would be the retrieval of copies of potentially useful documents by means of the Automatic Microimage File.

2. Magnetic Tape

The Special Purpose Electronic Searching Machine

One of the earliest suggestions for using a digital computer for information searching was made by Bagley. In 1951 Bagley and Perry presented a paper on this subject, based on Bagley's Master of Science thesis (270). The paper is interesting for two reasons. First, it gives Bagley's and Perry's early opinions on the division of labor between machine and manual searching. They state:

> Information must be analyzed and encoded in such a way that a search by machine for a given subject will direct attention to all the pertinent documents while rejecting nearly all those documents which are not of immediate interest. It is not advisable to require that machine searching do more than this and reject all material at or near the borderline of interest. In the first place, pinpoint accuracy in machine searching would require excessively detailed analysis of information. Furthermore, material near the borderline of interest may prove surprisingly useful to the person requesting the search. A reasonable division of labor should be the goal, with the machine screening out so much of the unwanted material that final review of the machine-selected items placees no excessive burdens on the person for whom the machine search was made (271).

The second interesting point made in this paper is the relative slowness of the general purpose digital computer when applied to information searching--this despite the fact that the then fastest computer, the Whirlwind I,

could perform 16,000 operations per second. Bagley and Perry write:

> Our study revealed that a search of average complexity, in which the machine is required to scan an encoded index...requires that Whirlwind in its present form would take about 3 seconds to scan and inspect the entries for a single document. By redesigning the computer element of the machine so that it might perform certain operations more simply than at present, it might be possible to reduce the time of searching of a single index block by a few tenths of a second--but no more. At this rate it would still take more than 800 hours to search the index to a million documents (272).

The reason for this slowness lies in the design of the general purpose digital computer. The machine has only one computing element to perform all the operations on the numbers in the computer. The rest of the machine is devoted to storage of numbers, to the control of these numbers when the machine is in operation, and to input and output devices. Much of the computer operating time is spent in transferring numbers back and forth and in keeping track of results of past operations. Bagley and Perry estimate that the actual identifying operations require 1% of the total running time; the other 90% is consumed, they believe, in transferring and storing intermediate results. As a solution to these problems, Bagley and Perry suggest the construction of a special purpose searching machine:

> A major advance in the rate of scanning and selecting could be accomplished if the machine were designed so that all criteria used to define the search would be checked simultaneously against each index entry being scanned. It would also be advantageous to design the electronic searching machine with a separate operations element to perform the two different types of operations--namely, one, detecting the identity of search criteria to index entries,

> and two, establishing the index entries pertaining to some one document stand in the specified relationship to each other. As a consequence, the special-purpose searching machine would be much simpler in design than the general-purpose computer. In particular, it is possible to design a relatively simple device--the comparator--for simultaneous scanning of index entries and another simple unit--the logical computer--for establishing that specified relationships exist between the index entries detected by the comparator (273).

The input into this machine would be the coded index as binary numbers on magnetic tape. In order to be able to run the tape through the machine at the rate of several hundred inches per second a special base for the magnetic tape is suggested. This is a durable alloy rolled into a strip of about 1 inch width. Programming the machine should eventually be by punched cards. A magnetic drum could be used for temporary storage of the output of the machine until it could be printed either by an electric typewriter or by a device which burns characters on a paper tape. The special purpose searching machine can be designed to scan the index for 1,333 documents every second. This is equivalent to 4.8 million documents per hour, though Bagley and Perry indicate that this speed might be reduced somewhat since information cannot be fed into the machine or removed from the machine at this rate.

The paper represents an interesting example of pure "blue sky" thinking. Although it is cited frequently in the documentation literature, no further work seems to have been done on the proposed system.

Standards Electronic Automatic Computer (SEAC)

The HAYSTAQ (Have You Stored Answers to Questions) system is an experimental machine literature searching system jointly developed by the U. S. Patent Office and the National Bureau of Standards. Subject matter selected for encoding into this system includes

chemical compounds, admixtures, and processes disclosed in patents. Searches are made on the SEAC computer. The SEAC is a general purpose digital computer located at the National Bureau of Standards offices in Washington. Changes have been or are being made in the computer's memory capacity, input medium, and logical operations repertoire to increase the computer's efficiency for this application. Another 512 words of mercury delay-line memory have been added bringing the total memory capacity to 2048 words. Eight Ampex tape units are being installed to meet the need for a large-capacity, high speed input medium. The disclosure file in the index will be stored on 6-channel tapes, which will be approximately 2,000 feet long with a packing density of 200 bits to the inch and a speed of 40-60 inches per second. A shift order for either right or left shift and an equality comparison order have been added to SEAC's command capabilities (274).

The final HAYSTAQ system will include:

1. A data preparation routine for the "library" making up the complete disclosure file of information;

2. A data preparation routine for the question;

3. The search routine with included subroutines;

4. The checkout routine which evaluates the apparent answers found to questions.

Only the search routine has been written and "debugged" (275).

Chemical disclosures--indexable information, in this case, to be encoded into the system--are grouped by document. The document is divided into compositions, that is, into physical admixtures of materials. A composition is further subdivided into ingredients or items. In the field of chemistry, descriptors are used

for functional groups, compound formulae, and molecular structure. The item itself is made up of a series of descriptor words. The levels of organization are thus document, composition, item, and descriptor.

In a search, the question is stored in SEAC's high speed memory and remains there throughout the search. The disclosure file, except for the encoded molecular structure data, is stored on a reel of magnetic tape called the principal tape. One disclosure composition at a time is read onto SEAC's memory and compared against pre-stored question compositions. The search progresses through various levels of organization. The question descriptor is matched against the disclosure descriptor. If no match is found for any question descriptor, further searching of the item is unnecessary and the search progresses to the next item. If any question cannot be matched, the search reverts again to the composition level and a new composition is inserted into the SEAC memory.

There are 3 types of descriptor searches:

Chemical descriptors (functional groups)
Empirical formulae
Molecular structure

When an apparent answer to a question item in the principal tape is found (a chemical descriptor or an empirical formula) a hit word is stored on one of the auxiliary magnetic tape units. If the question calls for a molecular structure search, the molecular structure data which is encoded onto a separate magnetic tape--called the secondary tape--is read into the SEAC. The principal and secondary tapes are so coordinated that coded structural data is immediately available when required (276).

In case all the question items in the composition have found answers the validity of the answers is checked by a machine program so that false drops can be eliminated. This is done by reading the complete set of hit words in from the tape and processing it through the check-out routine. In the case of a mixture

of materials the relationships among the hit words is checked to see whether these relationships are the same as the relationships sought in the question. When complete answers are checked out, the identification of the document is printed and the search continues with the next document (277).

The HAYSTAQ system requires an end to end search and various devices are used to speed up searches. The ordering of data is one such device. All descriptors in an item are arranged in ascending series according to their first digits which identify the category of subject matter. Whenever the first digit of a disclosure descriptor is greater than that of the question it is useless to look further in that disclosure item. The next item is considered. Data is also screened at the document level to eliminate fruitless searching. If, for example, the question requires a particular process, the document is examined to determine whether any process is included. If not, the entire document is eliminated from the search. Further screening is done to see if the document contains at least as many compositions as are required by the question.

An example of a chemical structure search called the topological structure search is given in an ORD report (278):

> Each atom (except non-significant H, which are the ones attached to the C of the hydrocarbon skeleton) and each significant bond (bonds other than single) in the structure is assigned an arbitrary serial number or 'interfix'. Consecutive numbers beginning with 1 are assigned to the bonds and elements in a completely random manner. The complete notation for each atom or bond coded is contained on one computer word and includes a symbol identifying it (e.g. by atomic number), its serial number, and the serial number assigned to each of the atoms or bonds to which it is connected. Both questions and disclosures are similarly coded. The serial numbers assigned to the bonds and elements in a question will seldom be the same as

> those assigned to a corresponding structural fragment of a disclosure compound containing the fragment sought. However, correspondence of numbers in this respect is not required by the computer program.... During the search, the machine inspects the first atom listed in the question code and each of the atoms in the disclosure code being considered until it finds a match. The atoms connected to this question atom are then determined and their correspondence to those connected to the first acceptable disclosure atom are compared until further matches are found. This continues until all of the bonds and atoms in the question, together with their proper connective relationships, are examined in the disclosure compound or until it is determined that no such total matching is possible. Note that the carrying forward of such comparisons is based upon identifications of elements and bonds and the identification of the adjacent elements and bonds.

Use: In 1957, a file of 250 complex steroid structural formulas was completely searched in 8 minutes. It was claimed that further experiments to determine the best method of asking a question have shown that it is possible to lower this time to approximately 8 seconds (279). In 1958, tests on the HAYSTAQ system were also conducted by simulating a computer on a blackboard (280).

HAYSTAQ is considered as an example of a general approach to a large scale, mechanized searching system. It is said to imitate to a large extent the search performed manually by a human searcher (281). The system is still considered highly experimental by its developers (282). It has been primarily an exercise in developing methodology in a field in which no guiding generalities have been established.

IBM 700 Series Electronic Data Processing Machines

The IBM 700 Series machines are general purpose large scale computers which, as their name indicates, are manufactured by the International Business Machines Corporation. The IBM 701 and 704 are primarily used for scientific and engineering calculation (considerable calculation on a relatively small amount of data), while the IBM 702 and 705 are used primarily for accounting or similar commercial applications (a small amount of calculation on a large amount of data.)

The IBM 704, for example, has the following features:

<u>Input</u>

Punched cards
Magnetic tape

<u>Memory</u>

8,192 words (10 decimals and a sign) magnetic core memory
8,192 or 16,384 words magnetic drum memory

<u>Output</u>

Off-line printer
On-line high speed printer
Cathode ray tube
Magnetic tape
Punched cards (283)

IBM 701:
U. S. Naval Ordnance Station, China Lake, California

The searching system at the China Lake U. S. Naval Ordnance Station includes only reports. It does not include periodicals and books. The subject matter is for the most part related to the development and testing of items of Naval ordnance (284). A manual Uniterm system used at this installation since 1953 was converted for IBM 701 searching although the manual Uniterm system itself was not discarded (285). When

the information was transferred onto tapes for machine searching 14,000 reports were covered by the system. About 9,600 descriptors were used and each document was indexed under about 8 descriptors. The index to 14,000 documents was stored on 1 1/3 reels of magnetic tape. About 300 new reports were added monthly (286).

Each descriptor was assigned a number and was punched on a set of IBM cards together with all the report numbers associated with that descriptor. Searching consists of an input, searching, and output phase. Bracken and Tillitt state:

> Phase I, Input: The input phase first read into memory up to 75 sets of punched cards. Each set of cards, called a data set, contains the necessary information to conduct one search. The first card of a data set gives the number of descriptors to be used in a search and controls the output printing of the search. For example, given eight descriptors for a search, it is possible to print the matching report numbers for all eight descriptors, plus the matching report number for the first two, the first three, through the first seven descriptors.
>
> Two additional cards complete a data set (only one card if six or less descriptors) and contain both descriptor numbers and particular report numbers. These descriptor numbers define the subject and the report numbers are used as starting points in the search.... The data sets are first written on a drum, the descriptor numbers for all data sets are then sorted (eliminating duplication) and the corresponding unit records on the library tape are found and written on tape....
>
> Phase II, Searching: Phase II executes the searching procedure using the working tape and the data sets as prepared by Phase I. A data set is first read from drum into memory. The first two unit records corresponding to the first two descriptors of the set are found

> on the working tape and stored in memory. The report numbers are then compared and matching ones saved. (They will be written on the result tape if a print, for these two alone, has been specified on the data set control card). The third unit card corresponding to the third descriptor of the data set will be read from the working tape and its report numbers previously found. This process is repeated for all of the remaining data sets.... Phase III, Output: Phase III prints the result tape. There may be a maximum of seven listings for each search if the search used eight descriptors, and if the data set control card specified a listing of matching report numbers from the first two, three...eight descriptors. For each listing, the identifying descriptors, the report numbers...and the matching report numbers are printed. If no matching report numbers are found then the word NONE will print (287).

A maximum of 8 descriptors can be coordinated in one search. Maximum space for 6, 750 report numbers has been provided for each descriptor (288). Sixteen library searches are made 3 times a week; a total time of 11 minutes is required for the 16 searches, not including the time it takes to change the tapes (289). Plans are being made to install new tapes having twice the density and twice the length of the present tape. A core memory will be added to the IBM 701. This is about twice as fast as electrostatic memory. Such improvements will cut down the time for 16 searches to approximately 5 minutes (290).

Bloomfield--formerly Head of the Acquisitions and Records Branch of the Library Division at the China Lake Naval Ordnance Station--comments on the first experiment with 3, 000 reports prepared for IBM 701 searching:

> I believed that the reference librarians could do all the coordinating necessary in the amount of time it would take for IBM processing....

> It is necessary to go to the accession list to find the titles of the accession numbers revealed by the Coordinate Index. This is also true for the accession numbers the IBM machine would produce. Also, one would have to go to the code to get appropriate numbers of the Uniterms so that the IBM computer could search its tape. If one needed to enlarge upon our initial approach by searching more Uniterms, the use of the Library's Coordinate Index would be simple, but if the IBM computer were involved it would mean rerunning the tape. Finally I was of the opinion that reference work requires some imagination and the IBM computer has none (291).

A note to Bloomfield's paper indicates that the paper does not represent the official view of the Naval Ordnance Test Station (292).

A note from Miss Canova, librarian at the Station, states that the searching program is scheduled to be transferred to the IBM 704 computer. She expresses the hope that results of the machine search can be expanded to give the corporate entry and title as well as the accession number (293).

The use of a high speed, higher cost computer in this installation can be questioned on several grounds. One was mentioned by Bloomfield when he indicated that it takes just as much overall time to do a search with a computer as it takes for manual searching of numbers. It must be remembered that in addition to the actual computer operating time, time must be added for translating descriptors into machine language and for programming the computer.

At present the computer search does not provide any more information than the equivalent manual search, namely document serial numbers of potentially pertinent documents. Plans are being made to include the corporate author and title as part of the search result, but it seems somewhat questionable whether this is more efficient than having a clerk manually pull the accession

cards of the documents themselves from the file.

Lastly, since the manual coordinate index is still being kept, it is doubtful whether the extra amoung of effort involved in preparing the computer based system is warranted even though "free time" might be available on the computer.

IBM 704 Electronic Data Processing Machine:
General Electric Company,
Aircraft Gas Turbine Division

Over 30,000 documents are included in this indexing system for scientists and engineers of the Aircraft Gas Turbine Division (AGT) of General Electric. The system is based on the use of descriptive keywords (descriptors) and document file numbers. All technical documents in the AGT library are identified in terms of these keywords. Each document may have a dozen or more words to describe it. There are over 7,000 keywords describing the documents. The arrangement is by document numbers on descriptor cards. The index, i.e. the document numbers on descriptor cards, and a concise abstract are entered on magnetic tape. The 30,000 abstracts are encoded on 3 reels of magnetic tape.

The retrieval system, using the IBM 704 computer, can search through the entire list of numbers in less than 3 minutes. The high speed printer can print out abstracts of documents found in the search in less than 15 minutes, depending on the number of documents found (294). It is stated:

> In its present form the system can accommodate 1,000,000 abstracts, 56,000,000 file numbers, and can perform up to 99 simultaneous literature searches.... New computing equipment is available which could increase the speed of our present system by 1,000 percent, while the document storage capacity would be around 10,000,000.... An automatic information retrieval system that searches out written

> information 1,000 times faster than a man can do has been developed and placed in operation for the technical library at General Electric's Aircraft Gas Turbine Division in Cincinnati, Ohio (295).

No other details are given about the system.

Evaluation: The IBM 704-based system, like the IBM 305 RAMAC-based system, employs an inverse order of filing. The filing of document numbers on descriptors here are merely listed. No relationships among descriptors are specified. It is not stated whether equal space is provided after each descriptor for document numbers. The method for entering document numbers on the descriptor cards is not indicated.

The system appears to be a mechanized version of the "Uniterm" or "Peek-a-Boo" system with one added feature. Abstracts of potentially pertinent documents can be printed out by the high speed IBM 704 printer. Whether this is more economical than having copies of these abstracts pulled out of the file and duplicated is not discussed.

How information is searched 1,000 times faster than with a manual system such as the manual "Uniterm" system equivalent is not mentioned.

IBM 705 Electronic Data Processing Machine:
Institute for Cooperative Research,
University of Pennsylvania

A system for searching document collections with an IBM 705 was designed at the Institute for Cooperative Research, University of Pennsylvania; bibliographic and subject information are to be encoded for machine searching.

> The documentary information fed into the computer memory is called the document record in this system. It consists of two sections: a bibliographic facts section and descriptor sec-

> tion. The bibliographic section contains seventeen facts about the document, such as identifying and locating number, the author or authors, the publisher, the date of publication, etc. The descriptor section of the document record contains an average of thirty descriptors per document to define the contents of the document (296).

The document is identified with a serial number. Information about one document is kept together and unless a division by subject or other dividing criteria is contemplated (this is not mentioned) the entire file has to be scanned for each search. Descriptors for each document are divided into arbitrary groups so as to reduce the incidence of false descriptor coordination, i. e. the coordination of unrelated descriptors in a document (297).

A special assembly of card punch equipment and a number of remotely controlled typewriters will be used to produce in one operation all the cards for the auxiliary card file and the necessary control records. The auxiliary card file is a simplified card catalog since the author, S. R. Moyer, feels that searches for certain types of information--e. g. searches for a particular document, author searches, and searches under broad classes--are still done more efficiently in the old-fashioned way (298). The necessary control records are accession records, the shelf list, and serials holdings.

Information to be searched by the IBM 705 is entered on magnetic tape from the punched card record. Each 2, 500 foot reel of tape will hold the bibliographic facts and descriptors of 14, 000 documents. The machine can search at the rate of 150, 000 documents per hour. The results of the search will be delivered by the IBM 705 printer in the form of a list of identifying and locating numbers for the documents. The printer can also print the bibliographic facts for each document (299).

The cost of the system is estimated. The rental

of the IBM 705 is $28,180.00 per month. The cost of personnel and material and the machine rental is estimated to be $41,180.00 per month.

> Under the system outlined, the machinery could search, barring breakdown and excluding multiple searching, 3.15 million documents every 24 hours (21 hours of work). If the machine were operated 20 days per month, the cost would be approximately $653.65 per search of 1,000,000 documents, or 18.9% of the $3450 necessary to carry out the same operation by a completely human agency (300).

Moyer bases the cost of a manual search on experience which he has had searching a collection of 1,000,000 documents.

The speed of the machine, 150,000 documents an hour, is compared with the speed of searching 6,897 documents per hour by "a completely human agency".

> Whereas the electronic computer actually searches every document record, the human agency searches only in certain areas of the card catalog where subject headings, etc., indicate that the document containing the required information are likely to be listed. Thus, vast areas of the catalog are eliminated without ever being touched (301).
>
> Claims:
>
> The IBM 705 system, staffed by the necessary crew, could do a literature search of a document collection in approximately 2% of man-hours required by a completely human agency; and the results of the search would be more accurate, thorough, and complete than the results of a search carried out by a human agency (302).

Use: The system is not yet in operation. No mention is made of any test with the system.

Evaluation: Several points made are open to question. Is a machine to replace the card catalog efficiently if it has to search through the equivalent of the entire card catalog for every search and if only one search can be done at one time? Is it bad that a human being can skip over large parts of the catalog which are of no interest to the search at hand? Are searches by machine automatically more accurate, thorough, and complete?

Univac

Univac I and Univac II are large scale, general purpose digital computers manufactured by the Sperry-Rand Corporation. Univac II, the later model, has the following features:

Input

Unityper, a modified electric typewriter which records characters directly on magnetic tape in binary code
Punched cards (through card-to-tape converter)
Magnetic tape

Memory

24,000 characters magnetic core memory

Output

Uniprinter, a modified electric typewriter actuated by magnetic tape signals
High speed on-line printer
Magnetic tape
Punched tape
Punched cards (303)

A study was made to determine the suitability of the Univac Fac-Tronic System to literature searching. The model studied consisted of a library of 1,000,000 documents. Each document was to be identified with an 8 digit serial number. The master reference file,

i.e. the index which was to be searched by the machine, would include the document serial number and other characteristic features of the document. These features might be: the author, date, contract number in the case of a government report, and subject or subjects. Each document would have an average of 15 characteristics encoded with a maximum of 30 characteristics (304).

The information would be encoded as numbers and letters. A 4 letter digit combination code would represent each item of information in a document. In an end to end search, the Univac is capable of searching a file of 1,000,000 documents averaging 15 approaches in approximately 4 hours. As the machine is presently constructed only one question can be answered at one time. By a simple modification of the circuitry to speed up the matching operation 6 to 10 questions can be answered in 4 hours. Schemes have been investigated that would decrease the search time to half an hour and increase the number of simultaneous searches by several fold. These modifications would, however, involve major changes in the design of the computer (305).

A later report on the possible application of the Univac for information retrieval was written by O'Connor (306). In this report 2 systems which would make use of the inverse ordering of document numbers on descriptors are discussed. An example of the speed of the system is indicated with a hypothetical index to 1,000,000 documents. A 6-descriptors search making use of logical sum, product and difference can be conducted in less than 5 and 12 minutes computer time respectively to select and print out the numbers of all the documents in the collection which satisfy that retrieval request (307). One of the systems requires several inexpensive modifications of the computer. For Univac II, a later model of the computer, all storage tape and search figures are cut by 50% or more (308).

Use: No further work was reported on the proposal by Mitchell; O'Connor's work seems to be in an early experimental stage.

Herner and Company Special Computer

A machine for literature searching is now being developed by Herner and Company, Washington, D. C. Herner states:

> The group has developed and built a small machine which records information in digital form on magnetic tape, searches the recorded information, performs simple correlations and indicates those entries that meet specific search conditions. Input is by keyboard; output is by solenoid-operated electric typewriter. Search conditions are specified by setting from one to three four digit numbers on rotary switches; the required correlation is specified by positioning another rotary switch. The machine then follows a fixed internally stored program of search, buffer and output operations. The basic construction of the machine is now completed, and it is being tested in the 'breadbox' stage before being mounted in a console. The capacity of the machine is such as to suit it for collections of 50,000 to 100,000 documents with an average of five coded entries each (309).

Lowry and Albrecht Special Computer

A special purpose computer for information retrieval and its possible application in a large organization is outlined by Lowry and Albrecht in a paper for the International Conference on Scientific Information (310). They state that:

> Information is available in amounts far greater than required by any single research organization and to the degree that it is extraneous to needs it constitutes a barrier to research progress if attempts are made to handle it in an information system (311).

Intelligent selection, a carefully defined and applied acquisition program, is essential in an information system. Even with such selection,

> in a large organization the amount of information stored will soon exceed the capacity of manual or common accepted retrieval methods (312).

A machine system to cope with such a situation and its possible application in large organizations is described in general terms:

> In most instances, the area of interest of a person seeking retrieval of information is not fully matched by the scope and setting of information stored and indexed for future reference. It is in this situation that the culling process achieved through the associations and correlations of the mind may be employed to advantage to determine what stored information may be of interest, even though it does not fully define the user's area of interest. A simple technique which approaches a function of the mind and which may be readily embodied in machine searching of information fields comprises the establishment of two classes of indexing terms, namely, general and essential (313).

This is similar to the technique used by the U. S. Patent Office in their work with the Bendix G--15--D computer. The technique is called weighting and is discussed above (314). The indexing term is specified as general or essential at the time a particular search is undertaken, not at the time of indexing.

In the proposed system index entries, which are not only single keywords but which can also be indexing phrases, names, etc., are entered with the document information on magnetic tape. The following housekeeping information is also included: document start and document end. The tape contains the index to the document in encoded form and is sent through the com-

puter during the search. The special purpose computer consists of the following parts:

1. A magnetic tape input device to feed the tape into the machine;

2. Storage blocks for identifier words (search terms). These can be included in the input device or can be a separate part of the machine;

3. Comparison circuits to match the indexing terms on the magnetic tape with the indexing terms in the storage block and to sense start and end of document signals;

4. Associated circuits programmed to detect essential and general terms and to send signals to a counter;

5. A counter to count the number of comparisons made and to send a signal to the output gate if the desired matches are made;

6. An output gate to signal that the document identity is to be read into the output device.

The output device can be a high speed direct printer, though this is not likely to be economical. Instead the document identity can be coded on another magnetic tape. This tape can be used to actuate a print out at the end of the search.

A filled out request form for a machine search is illustrated (315). The form contains all identifier words (descriptors) needed for the search in alphabetic and coded form. It also specifies whether each descriptor is general or essential and indicates the number of general descriptors that have to be matched.

Programming the machine is described only in general terms. It could be by setting keys and switches to establish the code for the search terms or it could be by punched card or magnetic tape programs. One block diagram shows a system with which only one

search can be done at a time (316). Another one depicts an expanded version which will permit 3 simultaneous searches.

In addition to using the system for retrospective searching, its use is suggested in a current information announcement system tailor-made for individual users' interests.

Use: The system is still in the machine development stage. A simulation of the plans proposed in the paper is contemplated on a simulation computer (317).

3. Magnetic Disc

IBM 305 RAMAC

The IBM 305 RAMAC has a large (5,000,000 alpha-numeric characters) random access memory which makes the machine potentially useful for inverse order (document number or descriptor) indexing systems. Two literature searching experiments with an IBM 305 RAMAC are reported in the literature.

One application of the machine to literature searching is at IBM's San José Research Laboratory. It was organized to gain practical experience in this area and to provide a functional service for the laboratory (318). Another IBM publication describes the operation of the machine (319).

> The IBM RAMAC...has a capacity of 5 million alpha-numeric characters. It consists of 50 aluminum discs 24 inches in diameter, mounted on a common shaft which revolves at the rate of 1200 r.p.m. Both sides of the discs are coated with an iron oxide layer upon which data may be recorded in the form of magnetized spots. The movable reading-recording arm contains two heads--one facing the top side of the disc and the other facing the bottom of the same disc. The arm may be moved vertically to any of the 50 discs and laterally to any one of the 100 concentric recording tracks (320).

In the experiment, the 5,000,000 character memory was divided into 50,000 record groups of 100 characters each. The memory was also divided into 3 units, with the following capacity in terms of record groups:

Dictionary: 1,000 record groups. Each dictionary 100 character record group contains up to 6 words and their home addresses. The 1,000 records of the dictionary will hold 6,000 words, each of which is an indexing term (descriptor) for the system;

Home records and overflows: 24,000 record groups. The home records are locations in the memory where document numbers corresponding to the indexing terms are stored. There are 6,000 home records, one for each indexing term and 18,000 overflows. Since each home record can only store 12 document serial numbers (according to Firth (321)--17 according to Nolan (322)--plus housekeeping information, additional storage has to be provided for the location of subsequent document numbers. These are stored in the overflow, along with the address of the home record for the same descriptor. Since space is not preassigned to each descriptor (beyond the first 12 serial numbers) considerable space is saved in the storage since there is no way of predicting how descriptors will be used and what the space requirements for each descriptor will be.

Bibliography: 25,000 records. One 100 character unit is used per indexed document and the following information is included: document number, library call number, author, date, and up to 56 characters of the title.

Firth (323) outlines the procedure used in the system. A sequential 5 digit serial number is assigned to each document and descriptors are selected for the document. Two punched cards are then prepared per document: one listing the descriptors and the document serial number, the other bibliographic information and the document serial number. Descriptors are limited to 10 characters because of the need for a fixed field length for machine reading. The 2 punched cards per document are then read into the RAMAC for storage in the memory.

In addition to the list of descriptors, phrase groupings can be created by linking related descriptors

(324). This is done by punching the code for a period between descriptors which should be linked. It is an instruction to the machine to assign a common 2 digit number in addition to the 5 digit serial number to the descriptors. Word sequence within the phrase can also be indicated by assigning a sequence digit in addition to the grouping digits and document serial number (325). This is intended to eliminate the "Venetian blind" versus "blind Venetian" type of false drop but can also be used to indicate the sequence of steps in a process as suggested by the Patent Office group (326).

In searching, the home addresses of the selected descriptors are typed on a punched card; they are then read into the RAMAC. Both logical sum and logical product searches can be made. Descriptors' home addresses are listed for logical product searches and are listed and separated by a comma for logical sum searches. According to Nolan:

> The search was put into operation as a result of information supplied to the machine by means of its own input--the punched card. No human intervention was necessary to set switches, or wire a control panel or modify a program for the particular search being processed (327).

Up to 10 terms can be matched simultaneously according to Nolan (328); up to 9 terms, according to Firth (329), who writes:

> Two basic programs and control panels are required for this application. The loading program which contains both dictionary and document loading requires 96 instructions. The search program is made up of 89 instructions and is applicable to any search without alteration. Programs are permanently stored in the file. This permits a complete change from one operation to another by merely changing the control panel and reading in a single card to withdraw the desired program and put it into operation (330).

Capacity of machine: A 5,000,000 character random access memory is very high for computers. Nolan points this out when he states that the maximum capacity of electronic computers is usually in the range of 50,000 to 100,000 characters (331). In the application illustrated by IBM, 12 serial numbers are posted on each record with an average of over 11 descriptors per document. An index to, and bibliographic information for, about 24,000 documents can be incorporated into the system (332).

Use: No information was given about the use of this particular file.

Another experimental application of the same machine is described in a Patent Office Research and Development report, Number 14 (333). The group of patents selected for this experiment are in the chemical polymer art. Subject matter includes properties, functions of inorganic and organic compounds, and processes involving these compounds. Three levels of descriptors are used: specific compounds (the most specific level) represented by a 3-digit code; specific structural fragments of these compounds, represented by a 2-digit code; and mutual attributes of the fragments, represented by a 1-digit code. The authors believe that more levels of search can be provided in order to encompass a more extensive or elaborate hierarchy (334).

All but the most generic level of terms are generated as needed from the actual terms used in the patents so that no pre-established dictionary is required for these terms. Interrelations among descriptors are indicated by adding one or more arbitrarily assigned digits at the end of the document numbers for pertinent descriptors. A computer program has been developed to recognize the symbols for the three levels of terms and their relationship. The computing operation is one of matching for logical product relationships and merging for logical sum relationships. Specifically, punched cards with the addresses of the portion of the files to be searched (descriptor addresses) and the logical grouping of the descriptors are inserted into the RAMAC. The computer will then seek out the sets of

data in its file corresponding to these addresses and perform a succession of merging, matching, and re-seeking operations until it arrives at the numbers of the documents satisfying the search requirements (335). The document numbers can then be printed out by the machine.

One of the yet unsolved problems of the system is mentioned:

> Where it is required to find a process containing A plus B and another process containing C plus D, it is not yet possible to avoid the invalid answer A plus B plus C plus D, all in the same process. Similarly, a fragment answering two separate sets of descriptors will respond as an answer to both (336).

Use: The system seems to represent one of a continuing set of experiments on the use of machines for searching operations at the U. S. Patent Office. No information is given on the use of the system.

Evaluation: The promising feature of the RAMAC is its large random access memory. This feature makes possible the use of the machine in an inverted order of storage type systems where relationships among descriptors may be indicated. The inverted order of storage eliminates the need for an end to end search. This was the principal advantage of the "Uniterm" and "Peek-a-Boo" systems, but something had to be traded to gain this advantage: the possibility of indicating relationships among descriptors in any but very elementary fashion. The RAMAC system permits both advantages, but at a cost. The rental price of $3, 000 plus per month is high, not when compared to the large general purpose computers but certainly when compared to the cost of catalogers, and the capacity of the machine is relatively low in terms of documents included in the system. In the IBM application the index to, and bibliographic data for, 24, 000 documents could be stored on one 5, 000, 000 character memory unit. If the bibliographic information were deleted, that is, if the answer to a search were only to be a document

number, the capacity of the machine would be doubled but this would still mean only 48,000 documents. Additional memory units can be incorporated into the system at a cost of $700 plus per month. It would seem that the RAMAC as it now stands is too costly for use in small collections and does not have the capacity for large collections.

III. Data Files

A. Dow Chemical Company

A file of coded chemical structures is stored on magnetic tape and searched with the IBM 704 at the Dow Chemical Company. Opler and Norton started development work on this project in 1952 (337). Punched card scanning equipment was considered impractical because complexities introduced by starting a search anywhere and continuing a search in any direction all involved requirements complicating the "untangling" of multiple substituents. Such substituents might be, for example, 4 chlorine atoms, 2 ortho to each other on a phenol and at least one of the others on a phenyl group attached to the phenol (338). This example illustrates a type of search where descriptors--in this particular case structural units--cannot merely be listed but have to be listed and the relationship among parts brought out.

In 1950, the chemical code consisted of 332 structure units (larger than a single atom but not as large as a "complex" group) (339). The structure units and the relationships among these units, both of which are reduced to a sequence of one and 3 digit numbers respectively, make up the code for chemical structures. The structure of a chemical is translated into groups of 7 digits which have the following meaning:

> The first digit gives the location in some other group (B) to which group (A) is attached to group (B). The second digit tells by which of its positions the group (A) is attached to group (B). The third, fourth, and fifth digits designate the group number (001-099) [the structure unit] assigned to group (A). The sixth digit is the identifying number of the group (B) to which

group (A) is attached. The seventh digit is the identifying number of the now coded group (A) (340).

Each chemical compound is identified by a serial number. The serial number and the molecular formula of each chemical compound are also encoded. The code is punched onto punched cards and converted to magnetic tape by means of a specially prepared program. Coded data is grouped in the following way on the tape:

Compound block: the code for the serial number, molecular formula, and structure of a chemical;
Block: 100 to 300 compound blocks or chemical compounds;
Reel: approximately 100 blocks.

Consequently, each reel of magnetic tape will contain 10,000 to 30,000 encoded chemicals (341).

A general search program has been prepared which, when modified by search request cards, can be used to handle 90% of the searches (342). The program consists of some 2,000 computer instructions, contained in a deck of approximately 100 binary punched cards (343). When searching for a particular structure, agreement is checked for on five criteria. These are:

1. Empirical formula;
2. Presence of structure units;
3. Indirect connection among structure units through a third structure group;
4. If #3 applies, tests are made for positional differences between points of attached pairs of structure units;
5. Direct connections between structural units.

If a compound fails to meet any of the criteria, it is rejected then and there, and another compound is searched (344).

The results of the search are serial numbers of selected chemicals. An alternative to this is the print

out of names of chemicals corresponding to the serial numbers. These are found by searching a special tape. A new technique has been developed in which the search results, that is, the chemical structures, are displayed pictorially on an oscilloscope. The structures are photographed so that they can be inspected at leisure (345). Chemical structures are made up of repeating patterns of letters, numbers, and simple geometrical designs. These are depicted as basic units including simple combinations of lines on a 64-dot square. The selected dots are thrown on the screen in succession but so quickly that the design appears instantaneously. Punched card codes consisting of 2-letter or number combinations were worked out to yield on demand any desired symbol in the 64-dot square (346).

A method for conducting up to 5 searches simultaneously, called multiplexing, has also been developed (347). In 1956, coding experience was obtained on 15,000 compounds (348). No later figures are given. Search speeds in the order of 10,000 structure comparisons a minute are obtained (349). A comparison was made between the accuracy of human and machine searching--disregarding speed. Searches for chemicals with certain specified structural features were made with the following results:

	Human	Machine
Correct retrieval		
Completely characterized	82	92
Partially indeterminate but known to be pertinent	15	15
Partially indeterminate with relative positions not known	--	26
Incorrect retrieval (fail-safe)	2	30
Failed to retrieve	13	5

From this experiment Opler concludes:

> Any mistakes the computer makes arise from human error in original coding or cause retrieval of compounds that do not fit the search criteria. No case of random failure has been detected to date.... We conclude that search

speed is reasonably high and accuracy is entirely satisfactory (350).

Use: No data is given on the use of the system.

B. Enjay Laboratories

At the Enjay Laboratories, butyl rubber compounding data is recorded on punched cards for searching on an IBM 101. Each tested compound is given an identifying serial number and data about the compound is recorded on 4 separate forms (351). The forms have been set up to correspond exactly to the layout of the IBM card so that the information can be directly punched onto the card. The individual forms contain the following information:

Cure and stress and strain data
Physical tests other than stress and strain data
Aging test
Electrical tests

The completed form sheets, called data sheets, are sent to the data processing clerk for card punching. One card is made out for each data sheet so that a maximum of 4 cards is made out per compound. The data sheets are microcarded and serve as the permanent records of the tests. Eight thousand cards, representing about 3, 500 compounds, were included in the file by 1957 (352).

The most important application of the file is that of searching for compounds possessing a combination of specified properties. The IBM 101 plug board is wired for the code of these properties. The individual decks (up to 4 if each type of property is being compared) are sorted through the machine at the rate of 450 cards per minute. The machine is programmed to print out serial numbers of pertinent compounds. Numbers on separate lists are matched manually.

Another application of the file is the preparation of lists of compounds arranged by ascending numerical

values of a particular property.

C. Federal Telecommunications Laboratories

The Electronic Spectroanalyzer developed by the Federal Telecommunications Laboratories combines a spectrophotometer and a digital computer (353). The instrument consists of 4 units:

A spectrophotometer to emit infrared rays and measure absorption
A recording device to encode the spectral data in numerical form on paper or magnetic tape
A reference "library" which contains the infrared absorptions of possible constituents in numerical form on digital tape
A high speed digital computer to do the mathematical calculations and record the analysis directly

The entire spectrum of a chemical mixture of up to 10 chemicals is analyzed in the spectrophotometer. The instrument records the mixture spectrum and converts the entire spectrum into numerical form on paper or magnetic tape. Each "library" spectrum is next compared to the unknown mixture by multiplying the "library" tape by a specimen tape at each wave length. The products are added together and used as coefficients in linear simultaneous equations. The answers confirm the presence or absence of constituents as well as the quantity of these constituents present (354). The form of the answer is not indicated.

The first commercial model of the Electronic Spectroanalyzer was scheduled to go to the Sloan-Kettering Institute in the fall of 1958 (355).

D. Midwest Research Institute

Data on chemical compounds is entered on magnetic tape for IBM 704 searching at the Midwest Research Institute in Kansas City, Missouri. The object-

ive of this program is the determination of new uses for chemical compounds by a correlation of chemical and physical characteristics with known uses (356). One tape is to be used for each of the following characteristics of chemicals:

Physical properties
Usage
Name
Chemical structure

Each chemical is identified by an arbitrarily assigned serial number on each of the tapes. Since there is only a total of approximately 1,875 chemical compounds in the system thus far, all 4 files are entered on one tape. During a search this data is transferred onto 4 working tapes (357).

The physical property data includes the empirical formula of the chemical compound and the physical constant for the melting point, boiling point, refractive index, density, and molecular weight. The physical constants are stored as a 2-decimal digit code. The empirical formula consists of the number of atoms in each of the 10 more commonly found elements in a special decimal-digit format.

Known uses for chemical compounds are coded as 6-decimal digit codes. Structure data is stored according to a modification of the code devised by Norton and Opler of the Dow Chemical Company. This code has been treated in an earlier part of the present section of this report.

Searches are programmed by punched cards. Ten sets of search criteria (10 descriptors from the physical property, usage, and structure unit categories) can be searched for simultaneously. Each set of test criteria is entered on one punched card as a series of 6-character words giving the category of the descriptor (physical property, usage, etc.) and the code for the descriptor as well as print out instructions (358). In a search for 10 descriptors, the machine can be programmed to list separately also chemical compounds

which satisfy less than 10 descriptors (359). An on-line printer--a printer that is directly connected with the computer--can be used for the output data; but Institute personnel feel that off-line printer operations are more efficient and hence make use of this mode of operation (360).

Use: The system was demonstrated at the International Conference on Scientific Information in Washington, D. C., in November, 1958, but no data has been given on its use.

E. Monsanto Chemical Company

Results of screening tests for chemicals, i. e. preliminary tests of the usefulness of a chemical for a particular application, are entered in a form which lends itself to searching and reproduction by an IBM 702 computer at the Monsanto Chemical Company.

A central data group receives all samples to be screened. This group assigns a 5 digit serial number to each sample and prepares the following records on punched cards:

Compound structure and identifying number
Compound molecular formula and identifying number (in later work the molecular formula record has been combined with the structure record) (361)
Compound name and identifying number (362)

The sample is then sent to the laboratory for testing. The test results are recorded by the scientists on notebook pages containing 80 columns. The results of a single test, including conclusions, usually occupy a single line. That is, they can be entered on an 80 column punched card. The first 12 columns are used for identification and contain the compound number, place of synthesis, date of testing, and field of application under test. Column 13 is used for one of the following general conclusions:

No further interest
Some activity, a lead for further study
Active, needs secondary screening or a field test (363)

The actual test results are also coded on some of the remaining columns. A carbon copy of the notebook page is sent to the tabulating department and the test result as well as the compound structure, compound name, and compound molecular formula, are entered on separated reels of magnetic tape.

The chemical structure code is a very simple one. It consists of only 20 structural units standing for chemical bonds and chemical elements--some single, some groups of elements. The encoding of a structure consists of rewriting the structure on cross-hatched paper. The elements and bonds are placed in the squares of the cross-hatched paper to compare exactly with the original chemist-written structure (364). The rules prepared so far are sufficient to handle the great majority of compounds, including many steroids. There is a relatively large number of compounds which require special handling.

The structure of each compound, the identifying number, and the molecular formula are punched onto punched cards. The first 5 columns of the card are punched with the compound identification number. Column 6 is a sequence control position for card numbering; the balance of the card is for structure data storage. A maximum of 10 cards may be used for each record, allowing storage of 740 characters of structure data. The instructions for the computer are stored permanently in a small deck of punched cards. All the data cards which are added to the structure file are entered in the computer along with this instruction deck. The group of cards which stand for the structure of a compound is transcribed through the computer onto a reel of magnetic tape in the form of a variable length record. Since the record size may vary from 45 to over 700 characters, the variable length feature allows maximum reading and writing speed in computer operations (365). The computer calculates the molecular

formula from the structure and inserts it in the proper place on the record (366). The computer also checks the accuracy of the input data by seeing, for example, that any atom does not have too few or too many bonds (367). One reel of tape may hold over 25,000 typical structure-molecular formula records (368).

In searching for a particular chemical structure, the exact structure specifications are transferred on IBM cards. The searching procedure is outlined by the authors:

> The program deck, containing all the instructions necessary to perform any type of search, is entered into the computer with control cards necessary to perform the specific search. We have available for high-speed scanning the molecular formula file, which is part of the structure record. As each structure is considered, the molecular formula is compared with the 'molecular formula' of the substructure or moiety sought. If the substructure contains strange elements, the structure being compared is rejected by the machine. Furthermore, if the search requirements demand a greater number of any given element than is presented in the particular stored structure, it will be rejected by the computer. If the desired molecular values are available in a structure, then the detailed search begins. With the aid of the control cards entered with the search program, the computer gets its ideas as to whether certain mutations of the substructure desired should be considered.... Depending on the nature of the search question, and control of information given the computer, the search continues until a match is found, or until the whole structure has been scanned, no equality found, and the structure rejected (369).

Something under 20,000 compounds are included in the system (370).

Use: In addition to making searches for chem-

ical structures, the system is used to prepare monthly summaries of screening reports (371). It is also used in searches for compounds tested under specified conditions (372).

F. Thermophysical Properties Research Center

The Thermophysical Properties Research Center (TPRC) is an industrially sponsored project established in 1957 to organize thermophysical data found in the literature, to conduct searches on this data, and to publish an annual volume of all the data in the files. About 20,000 references are to be coded annually (373).

The thermophysical properties and bibliographic citations to these properties are encoded for searching and processing on a Datatron Electronic Computer and auxiliary equipment (374). The Datatron is a medium-sized digital computer with a permanent magnetic drum memory of 4,000 machine words. A machine word in this particular case consists of 10 arithmetic digits and a plus or minus sign. In this installation the computer is equipped with 2 magnetic tape units, each with one tape. Each tape has a capacity of 400,000 machine words. Auxiliary equipment consists of an Electrodata 500 card-to-tape converter, IBM card punches, verifiers, sorters, reproducers, and tabulators (375).

Recorded information on thermophysical properties is selected from the following sources:

1. Abstract journals
2. Government and industrial research reports
3. Reports of private research institutions and universities, including theses
4. Major research centers throughout the world with which information exchange agreements have been established
5. Special collections, reference books, and compendia (376).

Thirteen items of information are used to characterize each reference. One item is a 7-digit identifying

serial number. Six items are for the bibliographic citation, one item is to indicate the language of the original document, and 5 items are used for the following subject information:

1. Properties: a 2-digit code field is used numerically to encode up to 99 properties;

2. Substance class: a 3-digit code field is used numerically to encode up to 999 substance classes;

3. Substance name: a 4-digit code field is used to encode up to 9, 999 substances within each class. The combination of substance class and substance name allows the encoding of a possible 100, 000, 000 names;

4. Physical state: one digit;

5. Type of subject coverage: one digit is used to indicate the nature of the treatment given in a reference (experimental, theoretical, property values, in various combinations) (377).

The encoded information for a reference is punched onto the first 40 columns of an IBM card.

The first 9 columns of the IBM card contain coded information on property, class, and substance. By sorting sets of punched cards which contain newly coded data by the first 9 columns, an ordered arrangement by property, by class within property, and by substance within class is obtained. If this set of data constitutes the initial storage for any one property of a substance, the information is stored in this order on magnetic tape. If the data represents new sets of information, i. e. information on the specific properties of a substance already in storage, the new data is stored in its proper, ordered location. A system of interfiling--sometimes called "banker sorting"--is used which maintains both a completely filled (packed) and ordered tape, thus making maximum use of available space on the tape.

> An empty tape is placed on one of the magnetic tape units and the active tape (the tape con-

> taining the previously stored information) on the other tape unit. The IBM cards containing the new information to be stored are read into the memory of the computer. The information from the active property tape is then copied into the previously empty tape until a location is reached where an insertion of new information is appropriate. This new information that has just been read into the memory of the computer is copied onto the previously empty tape. When all the appropriate insertions are made at this point, copying begins again from the original active tape. This process of alternate copying from tape unit to memory is continued until all new information has been interfiled in its proper location. The result is a new ordered and packed tape (378).

The file is searched by specifying the search requirements in punched card code form. The search requirements are entered into the computer and stored in its magnetic drum memory. The appropriate program tape is inserted into the computer and the retrieval program is activated. The area of the tape where the pertinent references are located (if the search can be so circumscribed) is located by means of an address directory at the head of the magnetic tape. This area of the tape or the entire tape is then matched against the search requirements which are stored in the computer's memory. When pertinent items are located, the serial number of the entry together with the other descriptive code numbers of the bibliographic citation are punched out on IBM cards by the output device of the computer. These cards are then fed into an IBM tabulator which prints the information coded on the punched cards (379).

In addition to searching for the thermophysical properties of a single chemical compound or a class of chemical compounds--the number of searches is not specified--the information contained in the system is compiled annually in printed form and sent to the system's subscribers (380). This information is issued in 3 parts. Part A is a classified list of substances with

code numbers of the properties listed for each substance and the code number of the substance. Part B is a list of properties, subarranged by classified substances in coded form. Also included in part B are the physical state, the type of subject coverage, the language, the year of publication, and the identifying serial number of the publication. Parts A and B are to be cumulated annually. Part C is a listing of references by identifying serial number and an author index to this list.

G. Electronic Structure Correlator
(No company association)

A special purpose computer has been proposed to search organic chemical structures. Organic chemical structures are translated into a notation called the Gordon-Kendall-Davison (GKD) notation, in which structural formulae are treated as topological systems of atoms connected by links, without considering bond multiplicities. Atoms are represented by their normal chemical symbols except that CH_3, CH_2 and CH groups are symbolized as J, L, and M respectively. All symbols are given an order of seniority. The senior atomic symbol is ciphered first; the senior symbol linked to the first cipher is linked next, and so on. A ring closure is indicated by an X (381).

The Electronic Structure Correlator which is to be used for the structure searching operation is a high speed, sequence-controlled electronic digital computer coupled to punched card reading and sorting equipment. The chemical structure is coded on punched cards. Each cipher is treated as a 12 digit binary number divided into 3 fields and occupying a single column on a standard punched card. One hole is punched if an atomic symbol is represented. A second field of 4 holes represents the coordination number (the number of ciphers to which this cipher is linked) and a third field of seven holes represents the atomic number. Carbon groups J, M, L are given pseudo atomic numbers (382). A link table--a table which lists the connections among ciphers--is prepared by machine from the structure

data in punched card form. This link table is stored in the computer's memory during a search. The machine scans the data for coexistence of structure elements (the ciphers) and then checks whether these ciphers are connected in a predetermined way.

The Electronic Structure Correlator will have 6 main memory units with a capacity of 80 symbols and 6 smaller memory units of 1-symbol capacity, together with 4 separate adders and the usual collection of "gates," "trigger controls," and auxiliary units. It will be considerably smaller than a large scale general purpose computer (383).

Use: There is no indication that any work has been done on the construction of any equipment mentioned in this paper.

IV. Miscellaneous Applications

A. Computer Preparation of Manual Coordinate Indices

Tabledex

A coordinate index which is to be partially prepared by a computer is suggested by Ledley (384). This coordinate index will have the following parts:

Part I: The bibliography proper, or list of articles with citations and, if desired, some descriptive material in addition to the citation.

Part II: An alphabetic list of descriptive words (descriptors) by means of which the retrieval is accomplished.

Part III: The indexing tables.

The tables which are illustrated below contain 2 types of numbers, the underlined article numbers down the left column and the non-underlined word numbers (descriptors) comprising the rows. There is one table for each distinct word of the word (descriptor) list and each table is numbered with this word number (385).

Table 5.1 (386)

<u>2.3</u>	5.2	5.3	7.1
<u>2.2</u>	5.2	6.1	7.1
<u>3.2</u>	5.2	6.2	7.1
<u>2.4</u>	5.3	6.2	
<u>3.4</u>	---		

The tables are used as in other non-manipulative coordinate indices. The word (descriptor) list is first checked for the identifying number of each descriptor. The table for the smallest number is then selected and the row which contains all the other given descriptor numbers is checked. The underlined article numbers for rows in which all the specified descriptor numbers are included represent articles of potential interest (387). Instead of translating the descriptors into numbers, the descriptors themselves can be used in the index tables (388).

The coordinate index preparation consists of the following steps. The indexer selects descriptors from each article. The bibliographic citation, additional information, and descriptors are translated into a form which a computer can handle, i. e. punched card and conversion to magnetic or paper tape. The computer will automatically assign both article numbers and descriptor numbers, if these are used, and will also form all tables. The output from the computer will be the completed bibliography, Parts I, II, and III, printed even with the correct page formats, ready for photo-offset duplication and binding into books. All this can be accomplished by the computer within a few hours (389).

No details of the computer program are given. There is no indication that any experimental work has been done on this system.

The size of a Tabledex for 10, 000 articles, 6, 000 words (3, 000 words with an average of 2 synonyms per word) and an average of 10 descriptors associated with each article, is estimated to be about 300 pages, if the print is similar to that of Webster's New Collegiate Dictionary, 1953 (390).

Univac

A computer-produced non-mechanized coordinate indexing system is proposed by O'Connor (391). The system is similar to that used by Batten (392), and this resemblance is acknowledged by calling the basic

units Batten cards (393). O'Connor suggests a card with a larger capacity than the one used by Batten. A card with 1,000 positions (100 columns and 10 rows) is proposed with the standard IBM card as a second alternative if the large card is impractical to mass produce (394). As in the Batten system, one card would be made out per descriptor. Each document would be identified by a serial number. The serial number would also stand for a position on the Batten (descriptor) card. This position would be punched on all Batten cards which applied to the indexed document. A search with this system would consist of selecting pertinent Batten cards, superimposing them and identifying the punched positions shared by all the selected Batten cards. The position numbers so identified would stand for serial numbers of potentially pertinent documents.

The computer-produced Batten cards, in this system, are supplemented with a computer-produced descriptive directory. This is an alphabetic list of descriptors, their code numbers (used instead of spelling them out, to save space on the Batten card), and the approximate number of documents entered under each descriptor. A document accession list is also part of the directory (394).

Both the Batten cards and the descriptive directory can be produced by a computer. Univac, for example, can produce magnetic tapes for a tape fed printer for the directory and for a tape-to-card converter for the Batten cards (395). Multiple copies can be made by gang-punching and print-copying devices respectively.

A hypothetical file is described to illustrate the speed of retrieval. This is a file of 100,000 documents which is indexed with a vocabulary of 5,000 descriptors. The index is entered on 50,000 punched cards with 1,000 one-line entries, each of 100 digit length. A search for 6 descriptors, making use of a logical sum, product, and difference, can be conducted in about one hour if each descriptor applies to about 1% of the collection and if the answer consists of about 20 documents. Searching the same file for a 4 des-

criptor logical product search requires only about 17 minutes (396).

The use of this system for the preparation of annual subject indices to technical journals is suggested. It could also be used to supplement mechanized systems so that the machine would only be used for the more difficult searches (397).

This work is still in the experimental stage.

B. Literary Data Processing

Preparation of Concordances by Computers

Data processing machines such as punched card sorters, computers and auxiliary equipment are now being used to assist the scholar in linguistic and conceptual analysis. Data processing techniques have been used for this purpose on the Summa Theologica of St. Thomas Aquinas (398, 399), part of the Dead Sea Scrolls (400, 401), and the Bible (402). In literary data processing the text is machine analyzed and indexed down to the simplest meaningful element, the word (403). The words are then compiled--with the assistance of the scholar--in a number of specialized lists such as the following:

1. An alphabetic list of all the words in the text as many times as they appear, along with the identifying data.

2. A list of all graphically different words (house and houses would appear separately in such a list), along with the number of times each word appears in the text and the alphabetic position which it occupies among all the words.

3. A list of all word families as defined by the scholar, who also examines all the different words and who groups the various forms of each graphic-semantic family under

a single parent expression or word. For example, the words were, are, be would be represented by the identity listing to be. Homographs are also separated in this list.

4. A lexicon which combines the features of the second and third lists. The lexicon will show the parent words, related component terms, frequency of each word's appearance, and the sum of frequency counts of all the words in the family.

5. A reverse index which combines the features of the first list and the third list. This list will show parent words, list related component words as many times as they appear, and show by reference number the location of each appearance by the word endings.

6. A concordance which will have for each word in the text one or several lines of text in which the word appears (404).

The machine procedure for the literary analysis of the Summa Theologica is outlined by Tasman:

1. The scholar analyzes the text, marking it with precise instructions for card punching.
2. A clerk copies the text using a special typewriter which operates a card punch. This typewriter has a keyboard similar to that of a conventional typewriter and produces the phrase cards. These cards contain all the lines or phrases of the text, one on each card, in sequence, transcribed in symbols (punched holes) that can be understood by the machine. Each phrase is preceded by the reference to the place where this line is found and provided by a serial number and a special reference sign. A second clerk types each phrase of the text on the appropriate phrase card which had already been punched using a checking machine. In this way the accuracy of the text cards is rigorously checked and cards containing trans-

cription errors are replaced.

3. From the phrase cards the machine automatically produces the word cards and a complete copy of the text, phrase by phrase. Each of these word cards contains only a single word of the text...accompanied automatically by identifying data (405).

The identifying data for each word includes:

a. The quotation of the place where the word is found.

b. The first letter of the preceding word and the first letter of the following word.

c. The number indicating the position which this word holds in the text, e.g. the 121st or 253rd word.

d. A special reference mark which characterizes the phrase to which the word belongs, e.g. "Here St. Thomas refers to another passage of his own work."

The context of the same word is printed on the reverse side of each word card. The horizontal space between punched lines is used and a maximum of 12 lines or an average of 100-120 words can be transcribed (406).

The machine prepares lists of different words on separate cards from the word cards. These cards are called form cards and contain in addition to the word the number of times it is found in the text and a number which indicates its alphabetic sequence in this list of words. This is done as follows:

a. The machine puts all word cards in alphabetic order.

b. The machine prints on sheets of paper the different words which it finds by examining all the word cards at a rate of about 6,000 cards per hour. It prints the first word. If the following word is different from it, it prints it. If it is identical, it does not print it, but counts it. After it has finished counting all identical words, it prints the total number next

> to the printed word. It proceeds then to print the following different words and so forth.... The machine punches at the same time another set of cards, one card for each different word (407).

The form cards are then analyzed by the scholar who separates homographs and combines words of one and the same graphic-semantic unit. The sum total of these words becomes the entry words. A card is punched for each entry word and is numbered sequentially. The form cards and word cards are then grouped under their respective entry card. A second code number (identifying the entry word) is automatically punched in all word and form cards. At the same time the machine adds to the entry card an identification of the total number of times in which it occurs in the text in one form or another. Lists such as those indicated can now be printed from these cards without any further effort on the part of the scholar. The printing speed ranges from 4,800 to 60,000 lines per hour depending on the type of machine used (408).

While conventional punched card machines are being used for the literary analysis of the Summa Theologica, a large scale computer, the IBM 705, is being used for such analysis of the Dead Sea Scrolls (409). The procedure appears to be similar to the all-punched-card technique except that the data is transcribed onto magnetic tape and processed more speedily.

The two principal advantages of machine technique compared with conventional literary analysis are greater speed and higher accuracy. The comparative speeds of the different techniques for indexing the approximately 13,000,000 of St. Thomas' complete words are estimated as follows:

Manual index: 50 scholars 40 man years	2,000 man years
Punched card machines: 10 scholars 4 man years	40 man years
Large scale data processing equipment, e.g. IBM 705: 10 scholars	

less than 1 man year 10 man years (410)

C. Autoabstracting

Autoabstracting is the selection by a machine of key sentences in a document and the reproduction of these sentences in their order of appearance in the original document. This work is described in 3 publications by Luhn (411), Savage (412), and Rowe (413), all workers at the International Business Machines Corporation.

In order to select the sentences in a document which are most suitable for the autoabstract, a measure of information count of each sentence has to be determined. This is called the "significance factor" of the sentence (414). The significance factor is based on the usage and position of significant smaller units in the sentence, that is, words. Significant words are determined by making a list of all of the words in a document and arranging them in descending order of use. This list of words is refined by eliminating types of words which have little discriminating power, e.g. conjunctions, articles, prepositions, and by combining words with common roots, e.g. differ, differentiate, different, differently, difference, differential, and considering these as one word. From this list of words significant words are chosen on the basis of frequency of usage in a particular document. In an illustrated autoabstract of a 2,326 word document, 39 words which occurred 5 or more times in the document were selected as significant words (415).

The position of these significant words in the sentences is next considered. Luhn's argument is:

> Whatever the topic, the closer certain words are associated, the more specifically an aspect of the subject is being treated. Therefore, wherever the greatest number of frequently occurring words are found in greatest physical proximity to each other, the probability is very high that the information being conveyed is

most representative of the article (416).

The significance factor of a sentence is derived from the number of significant words within the sentence and the linear distance between them due to the intervention of non-significant words, i.e. the proximity of significant words. An analysis of many documents has indicated that a useful limit is 4 or 5 non-significant words between significant words (417). All the sentences are ranked in order of significance factor and one or several of the highest ranking sentences may then be selected to serve as the autoabstract.

An experiment with autoabstracting is described by Savage (418). The document which is to be autoabstracted has first to be converted to a language which the machine understands. In the case of the IBM 704, this will have to be magnetized spots on magnetic tape. In order to save the time-consuming step of first translating the printed document word by word onto punched cards, a by-product of an earlier operation was used. This is the 31-channel punched paper tape which is prepared for the "Monotype" type casting operation. This tape was converted onto punched cards by means of a special tape and card converter. The cards were then converted onto magnetic tape by means of the standard card-to-magnetic-tape converter. This tape was fed into the IBM 704 for processing. The document is first separated into individual words and sentences, as 2 separate files. Common words are deleted from the list of words by means of a special look-up table. The remaining words are then alphabetized. A frequency study of the words is made; words with the same stem, e.g. different, difference, differently, are treated as one. In the sentence file, each sentence and the occurrence of each word are numbered serially from the beginning of the document. The average sentence length and the average word frequency are determined. A list of locations of any word which occurred more times than the average word frequency is produced. High frequency words are traced back to the sentences wherein they occurred and their position is noted. The portions of the sentences bounded by the high frequency words are then bracketed, providing that between any 2

high frequency words there are no more than 4 non-common low-frequency words. If this is not the case, 3 or more bracketed sections are constructed. The sentence is assigned a value corresponding to the square of the number of high frequency words, divided by the total number of non-common words in the bracketed section. The sentences are then sorted according to this value and the highest ones are selected for the autoabstract (419).

The method described has been used on about 50 articles, 300 to 4500 words each. The results have been encouraging enough to be further evaluated by a psychological experiment involving 100 people (420).

Luhn concludes:

> The results so far obtained for technical articles have indicated the feasibility of automatically selecting sentences that will indicate the general subject matter, very much as do conventional abstracts. What such autoabstracts might lack in sophistication they will more than compensate for by their uniformity of derivation (421).

D. Autoencoding

An indexing system in which the equivalent of indexing headings are selected by machine has been proposed by Luhn (422). The name which he assigns to this proposed system in another publication (423) is autoencoding. The words which serve the function of descriptors or index headings are called notions. Each subject area will have its own set of notions and as a preliminary step a set of notions peculiar to a particular field must be collected from an analysis of a representative sample of the collection. This job is done partially by machine and partially by human workers. The sample documents in a collection are transcribed into machinable form, i.e. onto punched cards, punched tape, or magnetic tape. Certain classes of words, e.g. nouns, and adjective qualifiers, are identified by special

symbols since they have to be recognized in subsequent steps of this procedure. These words or words with similar meanings will eventually become the notions. A card index of all the sentences is then prepared. A concordance of the words identified with special symbols is prepared from these cards. The next step is the grouping of words of similar or related meanings into notional families (424). This work is similar to that involved in the preparation of Roget's Thesaurus and the author suggests that the organization of the thesaurus serve as a basis for the operation (425). The preparation of the thesaurus, i.e. the organization of words into notional families, requires an expert knowledge of the subject. Individual words selected to characterize information should have about equal discriminating power as far as this particular body of information is concerned. The words--notions--will in this illustrative example be the nouns and adjective qualifiers, both of which were differentiated with special symbols. The notions are grouped into notional families, which in turn should have about equal discriminating power.

The final thesaurus will consist of 2 parts: the first part is the listing in some systematic order of the notional families, each identified by an index symbol such as a number or key word. Each of these would represent a listing of the words from the sample documents which are related with respect to the notion they express. The second part would be an alphabetic index of the words occurring in the first part, giving the key word or index number of the one or several notional families of which the given word is a member (426). The number of notional families is estimated to be less than 1,000 and this number will probably grow at a very low rate (427).

The autoencoding of the documents will be carried out with the aid of a dictionary of notions stored in the computer's memory. Each document is recorded in terms of notional elements and a pattern is thus created which will be used in searching for documents with related notions. A notion is assigned a special significance if it occurs twice in a paragraph or if it

occurs in succeeding paragraphs. Another way of selecting major notions--major notions is the name given to notions which have special significance--is to select words occurring in titles, headings, and resumés.

The machine operation is described by Luhn:

> All encoding would be carried out by a data-processing machine having a direct-access information storage and a look-up device. The dictionary index would be entered in one portion of the storage device and the document file previously prepared for making the concordance would now be processed in accordance with an encoding program. Each noun within a paragraph would be looked up and the corresponding family number or numbers extracted from the index storage and stored in a separate storage portion. The machine would then determine which of the words have a notion in common by comparing each family number in turn with the other numbers stored for nouns of the current, as well as the preceding and succeeding paragraphs. Since matching family numbers are indicative of a major notion, they would be entered in a third portion of the storage device. Words which fail to attain the status of major notions would not be entered.
>
> Upon exhaustion of this procedure and the recording of associated names, the encoded version of a paragraph would be ready for transfer to a permanent storage medium such as a reel of magnetic tape. Family numbers stored in the second portion of storage would be retained only during the analysis of the next-following paragraph. As the encoding of the new paragraph proceeds, the family numbers of its words would now also be compared with those of the immediately preceding and succeeding paragraphs to find common notions. This process would be repeated for each successive paragraph until the end of the chapter has been reached. The end result would be a

mechanically prepared notional abstract (428).

The collection of autoencoded documents would be searched by reversing the operation.

The inquirer will be asked to prepare an essay on what he wants, why he wants it, and anything else that might have a bearing on the question. The query document will then be translated into a notional abstract. This is done in the same way as in the case of the procedure used for encoding the documents in the collection. The query notional abstract is set up as the question pattern. The machine is asked to match this notion pattern with document notion patterns and to accept documents with a stated degree of similarity. An exact matching is unlikely in most cases since the answer would be an exact duplication of the question. To make the search more efficient, the search can be divided into 2 or more steps. The first step would be a rough screening for likely documents and this would eliminate the bulk of the file. The second step would be the search for the final specified conditions. If the search requirements have to be altered, subsequent searches would probably be done on a small part of the collection separated in the first screening operation instead of on the entire file of documents (429).

In a later article Luhn suggests the preparation of the autoabstract of a document and the index to the document (the autoencoded document) in one series of operations (430).

Use: The soundness and practicality of the system have not as yet been proven by any full scale operation. Experiment involving some 1,200 technical reports have produced encouraging results (431).

V. Summary and Conclusions

Machine based systems for literature searching, notwithstanding the vocal efforts by their proponents, have had very little impact on libraries or information centers. The type of usage which is made of information in the library and the overall economics of the problem seem to be the reasons for this. A large portion of the requests for information in libraries, perhaps the vast majority, seem to be of the reference question type which are answered relatively efficiently with the existing bibliographic apparatus and which would not be answered more efficiently with machine based systems. This type of request can be exemplified by a list of publications by an author, "a few good articles for background material on a subject," or the physical properties of a particular chemical. Reference questions such as these can be answered rather readily by the existing bibliographic apparatus which is the card catalog, published indices, reference books and other sources in that particular library or in other libraries, all coordinated by the reference librarian. This predominant kind of use of information in the library makes it necessary to keep the existing bibliographic organization and to superimpose on it something which can handle comprehensive searches more efficiently. Since these represent only a minority of the requests, particularly if we count only the comprehensive searches that cannot be handled efficiently by existing tools, the economics are against the installation of a machine based system in most libraries or information centers.

The reasons why existing methods are not efficient for some comprehensive searches are these:

1. The mere size of the literature in many fields makes the bibliographic apparatus too difficult to search.

2. Searches require different access to the information than that provided by the index. For example, a group of chemicals which are indexed by structural characteristics are difficult to search by type of use and vice versa.

3. The indexing is often not consistent or is incomplete. It should be pointed out that machine based systems will only speed up part of the searching operation because a part of the search does not lend itself to mechanization. This is pointed out by Shaw when he breaks down a search or bibliography, as he calls it, in the following steps:

I. Planning the bibliography: Determining the scope of the search, determining the sources to be searched, determining the headings under which the search will be made, estimating the time required and scheduling the operation.
II. Searching: Consulting the sources under the subjects indicated and selecting possible references; modifying the list of sources and of headings.
III. Copying the citations that appear to be pertinent.
IV. Locating copies of citations noted in sponsoring library and elsewhere.
V. Verifying the citations for accuracy and completeness.
VI. Analyzing the article to determine whether it is actually pertinent; preparing annotations or abstracts whenever this is called for.
VII. Organizing the material.
VIII. Editing the search report.
IX. Reproducing the search report in its final form (432).

Shaw points out that Steps I, VI, VII, and VIII, which represent a major portion of time in a high grade search, cannot be mechanized (433). Searching the index headings and copying the references by machine will result in the greatest savings. But scanning the index headings cannot be completely entrusted to the machine

for reasons outlined below. Consequently only part of the overall time for this operation will be saved and it is somewhat questionable whether in the average installation the savings in search time will compensate for the time and money invested in the machine and in indexing the information for machine searching if existing conventional indexing methods have to be maintained as well.

The machine can scan a large number of index entries more quickly and more efficiently than a human searcher if the search can be so phrased and programmed that the decision involved in accepting or rejecting a reference is a mechanical one. This qualification is emphasized because it is the fly in the ointment. No indexing systems, with the exception of an index to a subject which lends itself to unambiguous definition such as an index to chemical structure or to numeric data, is now at a stage where the scanning of entries is a complete mechanical operation. A good searcher does not trust the index heading alone; he knows that because of inconsistencies in the indexing, because of indexing from a different point of view than that of the search, because of indexing from the abstract instead of the original data, because the author did not spell out implied information for the indexer, information has to be read into the index heading and the abstract or additional information about the document has to be obtained before a decision is made to accept or reject the document. The danger here is not to select a borderline document but to reject documents on the basis of the index heading. In view of these deficiencies in an index, some of which can only be corrected by standardizing the authors, the scanning of index headings is not strictly a mechanical routine. Machines can be programmed to include a great deal of this borderline material but if this is done then the discriminating power of the machine search is greatly reduced and the bulk of the burden will again be placed on the human searcher. The literature on the use of machine based systems seems to substantiate these contentions.

Of the 42 electronic machine-based information re-

trieval systems discussed in the literature, 13 or only about 1/3 are in actual operation. The size and use of these systems, when size and use were indicated, seem to be low. Nine of these installations make use of the IBM 101. Twenty-four of the other reported systems are in the experimental stage, while the remaining 5 are in the "blue sky" stage.

Although fewer electronic machine based data files are discussed in the literature, a higher percentage of them are in actual operation. Of the 7 data files reported, 4 are in operation, 2 are in the experimental stage, and one is in the "blue sky" stage.

We should not assume from the present state of the art that the situation will remain static. There is no doubt in the writer's mind that additional work on indexing systems will not only improve the conventional manual systems but will also provide a more solid foundation for machine based indexing systems. This has been pointed out by Lowry:

> The present outlook for development of satisfactory machine techniques is very encouraging. This does not mean that we will have our machine system next year or even in five years but it is coming as surely as new cures for old maladies inevitably come. It will be forced upon us by demands of science and progress and solution may be found by those who need the information rather than by those who provide it (434).

In conclusion, we can say that the use of electronic machines in information retrieval systems is, to put it conservatively, not very widespread. This seems to be because machine based indexing systems in their present state of development have to supplement rather than replace existing conventional indices. The economics are against this in most situations. The picture looks brighter in areas where conventional indexing systems are not as deeply entrenched, or where the semantic problem is not as great. Data files seem to fall under this category. The use of machines to assist

the indexer in analyzing his vocabulary seems to be another promising application. Finally, the use of machines instead of human beings to do the intellectual tasks of abstracting and indexing is proposed as the ultimate solution. Like most ultimate solutions, this one seems to be a long way off.

References

1. Alexander, S. N. Machines for retrieving and correlating recorded information. (In: Shera, J. H., Kent, A., and Perry, J. W. Documentation in action. New York, Reinhold Publishing Corporation, 1956. p. 245-252). (p. 247).
2. Perry, J. W., Kent, A., and Berry, M. M. Machine literature searching. New York, Interscience publishers, 1956, 1956. (p. 51).
3. Reference 2, p. 52.
4. Reference 2, p. 57.
5. Berkeley, E. C., and Wainwright, L. Computers: their operations and applications. New York, Reinhold Publishing Corporation, 1956. (p. 29ff).
6. International Business Machines Corporation. IBM electronic statistical machine, type 101. New York, International Business Machines Corporation, 1953.
7. Reference 5.
8. Chapin, N. An introduction to automatic computers. New York, Van Nostrand Company, 1957.
9. Samain, J. Progres du classement et de la selection mecanique des documents: vers une memoire artificielle. Federation internationale de documentation. 17th conference, Berne, 1947. Rapports I. 1947. p. 22-26.
10. Samain, J. A new apparatus for classification and selection of documents and references by perforated cards. (In: Royal Society. The Royal Society scientific information conference, 1948, p. 680-685).
11. Metcalfe, J. Information indexing and subject cataloging. New York, Scarecrow press, 1957. (p. 187-190).
12. Jamieson, D. R. Mechanized bibliographical aid. Library association record 53:216-221 (1951). (p. 218).

13. Samain, J. Record card sorting device. U. S. Patent no. 2,618,386. Granted 18 November 1952. U. S. Patent Office. Official gazette 664:852 (1952).
14. Reference 9, p. 23-25.
15. Reference 10, p. 683.
16. Reference 10, p. 684-685.
17. Reference 10, p. 685.
18. Shaw, R. R. Mechanical storage, handling, retrieval and supply of information, Nato. Agard. Report 50 (1956). (p. 20).
19. Luhn, H. P. The IBM electronic information searching system. International Business Machines Corp., Yorktown Heights, New York, 1952. (11 p. and Appendix I-III).
20. Mechanized system launches new era for literature searching. Chemical and engineering news 30: 2806-2810 (1952).
21. Perry, J. W. Scientific aids for literature searching. Chemical and engineering news 30:3092 (1952).
22. New tools for the resurrection of knowledge. Chemical and engineering news 32:866-869, 891 (1954).
23. Reference 19, p. 3-4.
24. Reference 20, p. 2806.
25. Reference 20, p. 2808.
26. Reference 19, Appendix III.
27. Reference 22, p. 867-868.
28. Luhn, H. P. Card scanning mechanism. U. S. Patent no. 2,720,360. Granted 11 October 1955. U. S. Patent Office. Official gazette 699:293 (1955).
29. Reference 22, p. 868.
30. New punched card scanning equipment. (In: International federation for documentation. Manual on document reproduction and selection. 1953- . 746.12 E1-E2, April 1956). (p. E2).
31. Shaw, R. R. Machine literature searching. College and research libraries 17:450-452 (1956). (p. 451).
32. Himwich, W. A., Garfield, E., Field, G. G., Whittock, J. M., and Larkey, S. V. Final report on machine methods for information searching. Welch medical library indexing project, sponsored by the Armed Forces medical library.

Baltimore, Johns Hopkins University, 1955. (p. 5).
33. Taube, M. The Comac: an efficient punched card collating system for the storage and retrieval of information, AFOSR TN 58-365; Astia Doc. Ad. 154, 271 (October 1957). (p. 3-7).
34. U. S. Department of Commerce. Report to the Secretary of Commerce by the Advisory committee on application of machines to Patent Office operations, Washington, Government printing office, 1954. 76 p. (p. 57).
35. Reference 32, p. 6.
36. Reference 32, p. 6-7.
37. Reference 33.
38. Reference 33, p. 7.
39. Reference 33, p. 10-11.
40. Reference 33, p. 13.
41. Reference 33, p. 14.
42. Reference 33, p. 15.
43. Reference 33, p. 16-17.
44. National science foundation. Current research and development in scientific documentation. No. 2. 1958. p. 10.
45. Murphy, R. W. The IBM 9900 special index analyzer. Poughkeepsie, New York, International Business Machines Corporation, 1958.
46. Reference 45, p. 11.
47. Reference 45, p. 14.
48. Reference 33, p. 10-11.
49. Reference 45, p. 16.
50. Reference 45, p. 12.
51. Reference 45, p. 7.
52. Reference 45, p. 24.
53. Reference 33, p. 13.
54. Reference 33, p. 16-17.
55. Shaw, R. R. Machines and the bibliographical problems of the twentieth century. (In: Ridenour, L. N., Shaw, R. R., and Hill, A. G. Bibliography in an age of science. Urbana, University of Illinois press, 1951. p. 37-71). (p. 58).
56. Goldberg, E. Statistical machine. U. S. Patent no. 1,838,389. Application 5 April 1928, granted 29 December 1931. U. S. Patent Office. Official gazette 413:1217 (1931).

57. Green, J. C. The Rapid Selector--an automatic library. Revue de la documentation 17:66-68 (1950). (p. 66). (Reprinted from the Military engineer 41:350-352 (1949)).
58. Engineering Research Associates, Inc. Report for the microfilm Rapid Selector. (20 June 1949) Contract Cac-47-24. (PB #97313). 30 p. (p. iii, 2-3).
59. Shaw, R. R. The Rapid Selector. Journal of documentation 5:164-171 (1949). (p. 164).
60. Reference 58, p. iii.
61. The Rapid Selector. (July 1940). Cambridge, Massachusetts, Research Corporation, 1940. Unavailable, unverified reference.
62. Reference 58, p. 3.
63. Shaw, R. R. Mechanical storage, handling, retrieval, and supply of information. Nato. Agard. Report 50 (1956). (p. 21).
64. Reference 57, p. 66-67.
65. Reference 55, p. 59.
66. Reference 59, p. 164-165.
67. Photoelectric librarian. Electronics 22:122, 158, 160, 162, 164, 166 (September 19, 1949). (p. 122, 158).
68. Reference 57, p. 67.
69. Reference 58, p. 19.
70. Reference 58, p. 4-25.
71. Reference 67, p. 158-166.
72. Shaw, R. R. High speed intermittent camera. American documentation 1:194-196 (1950).
73. Reference 68, p. 4.
74. Reference 58, p. 25-26.
75. U. S. Department of Commerce. Report to the Secretary of Commerce by the Advisory committee on application of machines to Patent Office operations. Washington, Government printing office, 1954. 76 p. (p. 55).
76. Reference 58, p. 26.
77. Reference 55, p. 66.
78. Shaw, R. R. Management, machines, and the bibliographic problems of the twentieth century. (In: Chicago University. Graduate library school. Bibliographic organization. Chicago, University of Chicago press, 1951. p. 200-225). (p. 223).
79. Wise, C. S., and Perry, J. W. Multiple coding

and the Rapid Selector. American documentation 1:76-83 (1950).
80. Mooers, C. N. Coding, information retrieval, and the Rapid Selector. American documentation 1:225-229 (1950).
81. Wise, C. S. Multiple word coding vs. random coding for the Rapid Selector. A reply to Calvin N. Mooers. American documentation 3:223-225 (1952).
82. Optical punch card. Chemical industries 65:189-190 (1949).
83. Reference 75, p. 61-62.
84. Reference 63, p. 21-22.
85. Reference 63, p. 22.
86. National science foundation. Current research and development in scientific documentation. No. 2. 1958. p. 26.
87. Reference 55, p. 53.
88. Bedford, G. Review of current machine systems for handling information. Nato. Agard. Report 46. (1956). (p. 11).
89. Vickery, B. C. Some comments on mechanical selection. American documentation 2:102-107 (1951).
90. Shaw, R. R. System and apparatus for selective photographing. U. S. Patent no. 2, 594, 358. Application 24 April 1950, granted 29 April 1952. U. S. Patent Office. Official gazette 657:1346 (1952).
91. Steinhardt, L. R. Means for eliminating interference between the optical trains of a photographic reproducing apparatus. U. S. Patent no. 2, 621, 560. Application 17 November 1949, granted 16 December 1952. U. S. Patent Office. Official gazette 665:773 (1952).
92. Shaw, R. R. Photographic apparatus. U. S. Patent no. 2, 735, 335. Application 15 August 1952, granted 21 February 1956. U. S. Patent Office. Official gazette 703:494-495 (1956).
93. Engineering Research Associates. General description and proposed technical specifications for microfilm selector. Report. (August 1947). (Unavailable, unverified reference).
94. Engineering Research Associates. Progress report

for the microfilm Rapid Selector. Report. (November 1947). (Unavailable, unverified reference).
95. Butler. Instructions for the operation and maintenance of the microfilm Rapid Selector. Arlington, Virginia. Engineering Research Associates. n. d. 16 p. PB #97313s, supplement to ORS Cac-47-24. (Unavailable, unverified reference).
96. Bryant, M. S. Modern techniques of bibliographical production. Libri 4:143-152 (1954).
97. Engstrom, H. T. Microfilm selection equipment in information work. Industrial and engineering chemistry 42:1460-1461 (1950).
98. Gull, C. D. Instrumentation (in U.S. Government libraries). Library trends 2:103-126 (1953).
99. New machine will scan, select and copy research data at high speed. Library journal 73:797, 806 (1948).
100. Rapid Selector. Chemical and engineering news 27:1955 (1949).
101. Shaw, R. R. The Rapid Selector. AIBS bulletin 4:20-21 (July 19, 1954).
102. Shaw, R. R. Rapid Selector patented. Library journal 78:310 (1953).
103. Photographic selectors; the Rapid Selector and related devices. (In: International federation for documentation. Manual on document reproduction and selection. 1953- 752E1-E3, April 1956.
104. Samain, J. Filmorex; une nouvelle technique de classement et de selection des documents et des informations. (Paris, 1952). (Unavailable reference).
105. Filmorex; l'electronique au service de la pensee. (1955).
106. Samain, J. The organization of documentation by the Filmorex technique. Paris, Filmorex, 1956. 7 p. (Unavailable reference).
107. Filmorex system for electronic selection of microfilm cards. Monthly circular no. 7 for members of Unesco Committees and collaborating bodies concerned with documentation of the natural sciences, 5-6. Issued by Unesco, Paris. (15 December 1952). (Unavailable reference).
108. Samain, J. Une nouvelle technique de classe-

ment et de selection de documents. L'Onde electrique 36:671-675 (1956).
109. Samain, J. Documentation by the Filmorex technique. (In: Shera, J. H. Information systems in documentation. New York, Interscience publishers, 1957, p. 478-487).
110. Samain, J. Filmorex, eine neue technik elektronischer recherche für information und dokumentation. Nachrichten für dokumentation 9:35-40. (1958).
111. Filmorex. (In: International federation for documentation, Manual on document reproduction and selection. 1953- . 753E1-E2, April 1956.
112. Shaw, R. R. Mechanical storage, handling, retrieval and supply of information. Nato. Agard. Report 50. (1956). p. 22.
113. Reference 105, p. 2.
114. Reference 109, p. 481.
115. Reference 110, p. 40.
116. Reference 109, p. 483-484.
117. Reference 108, p. 672.
118. Reference 109, p. 483.
119. Reference 109, p. 484.
120. Reference 110, p. 38.
121. Reference 108, p. 673.
122. Reference 108, p. 674.
123. Reference 109, p. 486.
124. Reference 109, p. 478.
125. Coblans, H. New methods and techniques for the communication of knowledge. Unesco bulletin for libraries 11:154-175. (1957). p. 159.
126. Reference 110, p. 40.
127. Tyler, A. W., Myers, W. L., and Kuipers, J. W. The application of the Kodak Minicard system to problems of documentation. American documentation 6:18-30 (January 1955). (Also published in part in: International federation for documentation. Manual on document reproduction and selection. 1953- . 754E1-E4, 15 December 1957).
128. Kuipers, J. W., Tyler, A. W., and Meyers, W. L. A Minicard system for documentary information. (In: Shera, J. H. Information systems in documentation. New York, Interscience

publishers, 1957. p. 488-528). (Also published in American documentation 8:246-268. 1957).
129. Reference 127, p. 21.
130. Reference 127, p. 22.
131. Reference 127, p. 23.
132. Reference 128, p. 515-517.
133. Reference 128, p. 503.
134. Reference 127, p. 25.
135. Reference 127, p. 22-23.
136. Reference 127, p. 23-24.
137. Reference 128, p. 502.
138. Reference 128, p. 499.
139. Reference 127, p. 24.
140. Reference 127, p. 25.
141. Reference 128, p. 521-523.
142. Reference 128, p. 524-526.
143. Reference 128, p. 522.
144. Reference 128, p. 501.
145. Reference 127, p. 25-26.
146. Reference 127, p. 27.
147. Minicard demonstration. American documentation 6:258-259 (1955).
148. Lewis, C. M. and Offenhauser, W. H. Microrecording: industrial and library applications. New York, Interscience, 1958. (p. 61).
149. Reference 128, p. 488.
150. Hawken, W. R. Memorandum to M. A. Milczewski, University of California, 17 March 1958.
151. Shaw, R. R. Mechanical storage, handling, retrieval and supply of information. Nato. Agard. Report 50. (1956). (p. 23).
152. Taube, M. The Minicard system. (In: Taube, M. and associates. The mechanization of data retrieval. Washington, Documentation incorporated, 1957, p. 55-100).
153. Taube, M. and Heilprin, L. B. The relation of the size of the question to the work accomplished by a storage and retrieval system. AFOSR TN 57-483. Astia Ad No. 136476. 1957.
154. Bedford, G. M., et al. Final report on Phase I of the project Eye Bee - systems of information storage and retrieval. University of Pennsylvania, Philadelphia. 1 May 1955. Prepared under contract AF - 18(600) - 125. (Unavailable reference).

155. Bedford, G. M. Review of current machine systems for handling information. Nato. Agard. Report 46. (1956) (p. 11-12, 15).
156. Reference 155, p. 11.
157. Reference 155, p. 15.
158. Kent, A., Perry, J. W., and Egan, M. Minicard indexing study. First-Third reports. Western Reserve University. Astia Ad-101704, Astia Ad-96889, Astia Ad-91781. 1956.
159. U. S. Department of Commerce. Report to the Secretary of Commerce by the Advisory committee on application of machines to Patent Office operations. Washington, Government printing office. 1954. 76 p. (p. 57).
160. Reference 159, p. 57-58.
161. King, G. W., Brown, G. W., and Ridenour, L. N. Photographic techniques for information storage. Institute of radio engineers. Proceedings. 41:1421-1428 (1953).
162. Brown, G. W., King, G. W., and Ridenour, L. N. Considerations bearing on the use of photographic techniques for information storage. Mimeographed report. 26 March 1953. (Unavailable reference).
163. Reference 159, p. 56.
164. Introduction to the FLIP (Film library instantaneous presentation). American documentation 8: 330-331 (1957).
165. Avakian, E. A. Initial report on the automatic micro-film information system (AMFIS). Mimeographed report. 2 September 1952. Abstract: American documentation 4:174-175 (1953).
166. Avakian, E. A. Stored function calculator. U. S. Patent no. 2,610,791. Application 18 April 1950, granted 16 September 1952. Abstracted in U. S. Patent Office Official gazette 662:796 (1952).
167. The automatic micro-film information system (AMFIS). 1953. (Unavailable reference).
168. Lewis, C. M. and Offenhauser, W. H. Microrecordings; industrial and library applications. New York, Interscience, 1956. 456 p. (p. 289).
169. Avakian, E. A. and Garfield, E. AMFIS - the automatic microfilm information system. Special libraries 48:145-148 (April 1957).

170. Magnavox Company. Equipment for storage and retrieval. (In: National science foundation. Current research and development in scientific documentation. No. 2. 1958. p. 22).
171. Himwich, W. A., Garfield, E., Field, H. G., Whittock, J. M., and Larkey, S. V. Final report on machine methods for information searching. Welch medical library indexing project. Baltimore, Johns Hopkins University press, 1955. 38 p.
172. Bailey, M. F., Lanham, B. E., and Leibowitz, J. Problems of classification and documentation in the U. S. Patent Office in the field of petroleum and allied subjects. 3rd World petroleum congress proceedings, Section X, p. 13-21. Leiden, E. J. Brill, 1951.
173. Bailey, M. F., Lanham, B. E., and Leibowitz, J. Mechanized searching in the U. S. Patent Office. Journal of the Patent Office Society 35:566-587 (1953).
174. Kent, A. Non-conventional retrieval systems in documentation: preliminary comparative analysis. AFGSR TN, Technical note #3:58-575. June 24, 1958.
175. Edge, E. B., Fisher, N. G., and Bannister, L. A. System for indexing research reports using a punched card machine. American documentation 8:275-283 (1951) (p. 275).
176. Reference 175, p. 283.
177. Reference 175, p. 282.
178. Reference 175, p. 283.
179. Schultz, C. K. An application of random codes for literature searching. (In: Casey, R. S., Perry, J. W., and Berry, M. M. Punched cards, 2nd edition, New York, Reinhold, 1958. p. 232-247). (p. 232).
180. Reference 179, p. 232.
181. Reference 179, p. 243.
182. Reference 179, p. 239.
183. Reference 179, p. 244.
184. Reference 179, p. 245.
185. Schultz, C. K. Mechanized punched card systems for recording and searching the literature. Special libraries association, 43rd annual convention,

May 26-29, 1952. (p. 3).
186. Reference 185, p. 4.
187. Rockwell, H. E., Hayne, R. L., and Garfield, E. A unique system for rapid access to large volumes of pharmacological data; application to published literature on chlorpromazine. Federation proceedings 16: 726-731 (1957).
188. Reference 187, p. 730.
189. Reference 187, p. 727.
190. Reference 187, p. 727.
191. Reference 187, p. 730.
192. Hayne, R. L., and Turim, F. Machine retrieval of pharmacological data. (In: Shera, J. H., Kent, A., and Perry, J. W. Information systems in documentation, advances in documentation and library science II. New York, Interscience publishers, 1957. p. 240-249). (p. 246).
193. Reference 187, p. 729.
194. Reference 192, p. 247.
195. Reference 187, p. 730.
196. Reference 187, p. 730.
197. Crandall, G. S., and Stumpf, P. Q. A corporation's international network: an experimental demonstration. (In: Shera, J. H., Kent, A., and Perry, J. W. Information systems in documentation. Advances in documentation and library science II. New York, Interscience publishers, 1957. p. 572-583). (p. 579).
198. Socony Mobil Oil Company. Research and development laboratory. Machine literature processing demonstration manual. Paulsboro, Socony Mobil Oil Company, April 1, 1957. (p. 32).
199. Reference 198, p. 31.
200. Reference 198, p. 30.
201. Reference 197, p. 575.
202. Reference 197, p. 579.
203. Reference 198, p. 22.
204. Reference 198, p. 32.
205. Crandall, G. S., and Brown, B. M. An information service using both hand and machine sorted punched cards. Journal of chemical education 25: 195-199, 203 (1948).
206. Jones, W. S., and Butterfield, P. H. A technical information service using punched cards for in-

dexing and retrieval. (In: Peakes, Gilbert L., Kent, A., and Perry, J. W. Progress report in chemical literature retrieval. Advances in documentation and library science I. New York, Interscience publishers, 1957. p. 57-68). (p. 58).
207. Reference 206, p. 62.
208. Reference 206, p. 58.
209. Luhn, H. P. Information retrieval through row by row scanning on the IBM 101 electronic statistical machine. Yorktown Heights, N. Y., International Business Machines Corporation, November 17, 1958.
210. Reference 209, p. 2.
211. Reference 209, p. 2.
212. Reference 209, p. 4.
213. Reference 209, p. 2.
214. Luhn, H. P. The IBM universal card scanner for punched card information searching systems. Yorktown Heights, N. Y., International Business Machines Corporation, November 17, 1958.
215. Reference 214, p. 1.
216. Reference 214, p. 2.
217. Reference 214, p. 2.
218. Reference 214, p. 3.
219. Reference 214, p. 4.
220. Reference 214, p. 13.
221. Reference 214, p. 8.
222. Reference 214, p. 9.
223. Reference 214, p. 14.
224. Reference 214, p. 14.
225. Reference 214, p. 22.
226. Murphy, R. W., International Business Machines Corporation, personal conversation November 18, 1958.
227. Reference 214, p. 14.
228. Samain, J. La Mecano memoire. Hommes et techniques 23/24:59-61 (1945).
229. Bailey, M. F., Lanham, B. E., and Leibowitz, J. Mechanized searching in the U. S. Patent Office. Journal of the Patent Office Society 35: 566-587 (1953).
230. Perry, J. W., Kent, A., and Berry, M. M. Mechanized literature searching--a progress report. College and research libraries 16:162

(1955).
231. Andrews, D. D. Interrelated logic accumulating scanner (ILAS). U. S. Patent Office. Research and development report #6. 1957. (p. 3).
232. Reference 231, p. 5.
233. Perry, J. W., Kent, A., and Berry, M. M. Machine literature searching. New York, Interscience publishers, 1956. (p. 83).
234. Reference 231, p. 5.
235. Lanham, B. E., Leibowitz, J., and Koller, H.R. Advances in mechanization of patent searching--chemical field. Preliminary report, Office of research and development, U. S. Patent Office, April 11, 1956. (p. 10).
236. Reference 231, p. 8.
237. Reference 231, p. 9.
238. Reference 231, p. 10.
239. Reference 231, p. 12.
240. Reference 231, p. 11.
241. Frome, J., Koller, H. R., Leibowitz, J., Pfeffer, H., and Andrews, D. D. Recent advances in Patent Office searching: Steroid compounds and ILAS. U. S. Patent Office, Research and development report #8, 1957. (p. 14).
242. Reference 241, p. 15.
243. Frome, J., Leibowitz, J., Andrews, D. D., Grandine, J. D., Polyak, S. T., and Siedschlag, K. G. A system of retrieval compounds, compositions, processes, and polymers. U. S. Patent Office, Research and development report #13, 1958.
244. Leibowitz, J., Frome, J., and Andrews, D. D. Variable scope search system: VS_3. International conference on scientific information, preprints, area V. Washington, D. C., International conference on scientific information, 1958. p. 291-316.
245. Reference 231, p. 5.
246. Reference 244, p. 304.
247. Reference 244, p. 305.
248. Reference 241, p. 15.
249. Reference 231, p. 12.
250. Kent, A., Melton, J., and Flagg, C. Abstracting, coding, and searching the metallurgical literature

for A. S. M. The WRU searching selector. (In: Shera, J. H., Kent, A., and Perry, J. W. Information systems in documentation. Advances in documentation and library science II. New York, Interscience publishers, 1957. p. 407-434). (p. 409).
251. Perry, J. W., Kent, A., and Berry, M. M. Machine literature searching. New York, Interscience publishers, 1956.
252. Reference 251, p. 78.
253. Reference 251, p. 84.
254. Reference 251, p. 86.
255. Reference 251, p. 86.
256. Reference 250.
257. Reference 250, p. 431.
258. Search selector makes debut. Chemical and engineering news 35:116 (May 13, 1957).
259. Reference 250, p. 429.
260. Shaw, R. R. Review of Perry, J. W., Kent, A., and Berry, M. M. Machine literature searching ... College and research libraries 17: 450-452 (1956).
261. Reference 260, p. 451.
262. Reference 250, p. 421.
263. Weik, M. H. A second survey of domestic electronic digital computing systems. Report #1010, U. S. Aberdeen Proving Ground, Maryland, June, 1957.. p. 32.
264. Frome, J., Leibowitz, J., Andrews, D. D., Grandine, J. D., Polyak, S. T., and Siedschlag, K. G. A system of retrieval compounds, compositions, processes, and polymers. U. S. Patent Office, Research and development report #13, 1958.
265. Reference 264, p. 7.
266. Reference 264, p. 8.
267. Automatic microimage file. Computers and automation 5: 26-27 (1956).
268. An automatic microimage file searches film record and makes photographic print. National bureau of standards Technical news bulletin 40: 89-90 (1956).
269. Reference 268, p. 89-90.
270. Bagley, P. R., and Perry, J. W. Applicability

of newer electronic techniques to information searching. American chemical society, Division of chemical literature. New York, 1951.
271. Reference 270, p. 1.
272. Reference 270, p. 3.
273. Reference 270, p. 4.
274. Pfeffer, H., Koller, H. R., and Morden, E. C. A first approach to patent searching on standards electronic computer (SEAC). U. S. Patent Office Research and development report #10, 1958. (p. 1).
275. Reference 274, p. 2.
276. Reference 274, p. 5.
277. Reference 274, p. 16.
278. Lanham, B. E., Leibowitz, J., Koller, H. R., and Pfeffer, H. Organization of chemical disclosures for mechanized retrieval. U. S. Patent Office, Research and development report #5, 1957. (p. 9).
279. Reference 278, p. 10.
280. Koller, H. R., Morden, E. C., and Pfeffer, H. The Haystaq System: past, present, and future. International conference on scientific information, Preprints, Area V. Washington, D. C., International conference on scientific information, 1958. p. 317-353. (p. 346).
281. Reference 280, p. 318.
282. Reference 280, p. 340.
283. Weik, M. H. A second survey of domestic electronic digital computing systems. Report #1010, U. S. Aberdeen Proving Ground, Maryland, June, 1957. (p. 164-168).
284. Bracken, R. H., and Tillitt, H. E. Information searching with the 701 calculator. Journal of the Association for computing machinery 4:131-136 (1957) (p. 131).
285. Bloomfield, M. Evaluation of coordinate indexing at the Naval Ordnance Station. American documentation 8:22-25 (1957) (p. 24).
286. Reference 284, p. 132.
287. Reference 284, p. 133.
288. Reference 284, p. 133.
289. Reference 284, p. 135.
290. Reference 284, p. 135.

291. Reference 285, p. 24.
292. Reference 285, p. 22.
293. Canova, M. F., letter to M. J. Voigt, April 8, 1958.
294. Dennis, B. K. Rapid retrieval of information. Computers and automation 7:8-9 (1958). (p. 8).
295. Reference 294, p. 8.
296. Moyer, S. R. Automatic search of library documents. Computers and automation 6:24-29 (1957) (p. 27).
297. Reference 296, p. 27.
298. Reference 296, p. 26.
299. Reference 296, p. 27.
300. Reference 296, p. 28.
301. Reference 296, p. 29.
302. Reference 296, p. 29.
303. Weik, M. H. A second survey of domestic electronic digital computing systems. Report #1010, U. S. Aberdeen Proving Ground, Maryland, June, 1957. (p. 336-337).
304. Mitchell, M. F. Use of Univac Fac-Tronic system in the library reference field. American documentation 4: 16-17 (1953). (p. 16).
305. Reference 304, p. 17.
306. O'Connor, J. J. Information retrieval by Univac and Univac-produced non-mechanized systems. Part I. Univac applications research center, Technical report #18, 1957.
307. Reference 306, p. 3.
308. Reference 306, p. 4.
309. Herner and Company. Equipment for storage and retrieval. (In: National science foundation. Current research and development in scientific documentation #2. Washington, D. C., 1958. (p. 5).
310. Lowry, W. K., and Albrecht, J. C. A proposed information handling system for a large organization. International conference on scientific information, Preprints, Area V. Washington, D. C., International conference on scientific information, 1958, p. 355-376.
311. Reference 310, p. 356.
312. Reference 310, p. 362.
313. Reference 310, p. 363.
314. Frome, J., Leibowitz, J., Andrews, D. D.,

Grandine, J. D., Polyak, S. T., and Siedschlag, K. G. A system of retrieval compounds, compositions, processes, and polymers. U. S. Patent Office, Research and development report #13, 1958.
315. Reference 310, p. 365.
316. Reference 310, p. 366.
317. Lowry, K. W., personal conversation, February 3, 1958.
318. Firth, F. E. An experiment in literature searching with the IBM 305 RAMAC. San Jose, California, International Business Machines Corporation, November 17, 1958. (p. 1).
319. Nolan, J. J. Principles of information storage and retrieval using a large scale random access memory. San Jose, California, International Business Machines Corporation, November 17, 1958.
320. Reference 319, p. 1.
321. Reference 318, p. 5.
322. Reference 319, p. 6.
323. Reference 318, p. 1.
324. Reference 318, p. 7.
325. Reference 319, p. 11.
326. Lanham, B. E., Leibowitz, J., and Koller, H.R. Advances in mechanization of patent searching--chemical field. U. S. Patent Office, Research and development preliminary report, April 11, 1956. (p. 7).
327. Reference 319, p. 9.
328. Reference 319, p. 9.
329. Reference 318, p. 9.
330. Reference 318, p. 10.
331. Reference 319, p. 1.
332. Reference 318, p. 10.
333. Leibowitz, J., Frome, J., and Andrews, D. D. Variable scope searching by an inverted file technique. U. S. Patent Office, Research and development report #14, 1958.
334. Reference 333, p. 7.
335. Reference 333, p. 6.
336. Reference 333, p. 6.
337. Opler, A. Dow refines structural searching. Chemical and engineering news 35:92-96 (1957).

(p. 92).
338. Opler, A., and Norton, T. R. New speed for structural searches. Chemical and engineering news 34: 2812-2814, 2816 (1956). (p. 2814).
339. Reference 338, p. 2812.
340. Norton, T. R., and Opler, A. A manual for coding organic compounds for use with a mechanized searching system. Midland, Michigan, Dow Chemical Company, 1956. (p. 1).
341. Opler, A., and Norton, T. R. A manual for programming computers for use with a mechanized system for searching organic compounds. Pittsburg, California, Dow Chemical Company, 1956. (p. 2).
342. Reference 337, p. 96.
343. Reference 337, p. 93.
344. Reference 341, p. 4.
345. Chemical structures by computer. Chemical and engineering news 36: 108-109 (1958). (p. 108).
346. Reference 345, p. 108.
347. Reference 337, p. 93.
348. Reference 338, p. 2813.
349. Opler, A., and Baird, N. Experience in developing information retrieval systems on large electronic computers. International conference on scientific information, Preprints, Area IV. Washington, D. C., International conference on scientific information, 1958. p. 37-48. (p. 42).
350. Reference 337, p. 95.
351. Neu, R. F. Electronically sorted punched cards for polymer compounding data. Rubber world 137:84-90 (1957). (p. 85).
352. Reference 351, p. 86.
353. Chemicals spectra taped. Chemical and engineering news 36:62 (1958). (p. 62).
354. Reference 353, p. 62.
355. Reference 353, p. 62.
356. Findley, L. D., Bolze, C. C., and Carpenter, R. A. A card-controlled routine for searching chemical compound data with an IBM 704. Kansas City, Missouri, Midwest Research Institute, 1958. (p. 1).
357. Reference 356, p. 2.
358. Reference 356, p. 7.

359. Reference 356, p. 11.
360. Reference 356, p. 1.
361. Waldo, W. H., and De Backer, M. Printing chemical structure electronically: encoded compounds searched generically with IBM 702. International conference on scientific information, Preprints, Area IV. Washington, D. C., International conference on scientific information, 1958, p. 49-68. (p. 60).
362. Waldo, W. H., Gordon, R. S., and Porter, J. D. Routine report writing by computer. American documentation 9: 28-31 (1958). (p. 29).
363. Reference 362, p. 29.
364. Reference 361, p. 54.
365. Reference 361, p. 57.
366. Reference 361, p. 58.
367. Reference 361, p. 58.
368. Reference 361, p. 60.
369. Reference 361, p. 61.
370. Reference 361, p. 67.
371. Reference 362, p. 31.
372. Reference 361, p. 66.
373. Touloukian, Y. S., Stevens, C. H., Rodine, R. H., Wing, T., and Smith, D. W. Systems and procedures developed for the search and coding and mechanized processing of bibliographic information on thermophysical properties. Preprint of paper to be presented at the ASME symposium on thermal properties, Purdue University, February, 23-26, 1959. (p. 16).
374. Reference 373, p. 11.
375. Reference 373, p. 11.
376. Reference 373, p. 2.
377. Reference 373, p. 4.
378. Reference 373, p. 13.
379. Reference 373, p. 15.
380. Reference 373, p. 16.
381. Davison, W. H., and Gordon, M. Sorting for chemical groups using Gordon-Kendall-Davison ciphers. American documentation 8: 202-210 (1957). (p. 204).
382. Reference 381, p. 205.
383. Reference 381, p. 208.
384. Ledley, R. S. Tabledex: a new coordinate index-

ing method for bound book form bibliographies. International conference on scientific information, Preprints, Area V. Washington, D. C., International conference on scientific information, 1958. (p. 394-417).
385. Reference 384, p. 398.
386. Reference 384, p. 408.
387. Reference 384, p. 398.
388. Reference 384, p. 413.
389. Reference 384, p. 402.
390. Reference 384, p. 397.
391. O'Connor, J. J. Information retrieval by Univac and by Univac-produced non-mechanized systems. Part I. Univac applications research center, Technical report #18, 1957.
392. Batten, W. E. Punched card systems of indexing to meet special requirements. Report of the proceedings of the 22nd ASLIB conference. London, ASLIB, 1947. (p. 37-39).
393. Reference 391, p. 76.
394. Reference 391, p. 79.
395. Reference 391, p. 73.
396. Reference 391, p. 4.
397. Reference 391, p. 5.
398. Busa, R. The use of punched cards in linguistic analysis. (In: Casey, R. S., Perry, J. W., and Berry, M. M. Punched cards, 2nd edition. New York, Reinhold, 1958. p. 357-373).
399. Tasman, P. Literary data processing. IBM journal of research 1: 249-256 (1957).
400. Tasman, P. Indexing the Dead Sea scrolls by electronic literary data processing methods. New York, International Business Machines World Trade Corporation, 1958.
401. Secrest, B. W. The IBM 101 electronic statistical machine applied to an analysis of the Dead Sea scrolls. New York, International Business Machines World Trade Corporation, 1958.
402. McCulley, W. R. Univac compiles a complete Bible concordance. Systems 20:22-23 (1956). (Reference not verified).
403. Reference 399, p. 249.
404. Reference 400, p. 6.
405. Reference 399, p. 253.

406. Reference 399, p. 254.
407. Reference 399, p. 255.
408. Reference 399, p. 255.
409. Reference 400, p. 3.
410. Reference 400, p. 11.
411. Luhn, H. P. Automatic creation of literature abstracts. IBM journal of research 2: 156-165 (1958).
412. Savage, T. R. The preparation of autoabstracts on the IBM 704 data processing system. Yorktown Heights, International Business Machines Corporation, 1958.
413. Rowe, H. T. Scientific abstracting by computers. Computers and automation 7: 12 (1958).
414. Reference 411, p. 160.
415. Reference 411, p. 163.
416. Reference 411, p. 160.
417. Reference 411, p. 160.
418. Reference 412.
419. Reference 412, p. 4.
420. Reference 411, p. 162.
421. Reference 411, p. 165.
422. Luhn, H. P. A statistical approach to mechanized literature searching. IBM journal of research 1: 309-317 (1957).
423. Luhn, H. P. A business intelligence service. IBM journal of research 2: 314-319 (1958).
424. Reference 422, p. 314.
425. Reference 422, p. 314.
426. Reference 422, p. 314.
427. Reference 422, p. 315.
428. Reference 422, p. 315.
429. Reference 422, p. 316.
430. Reference 423, p. 316.
431. Reference 422, p. 316.
432. Shaw, R. R. Mechanical electronic aids for bibliography. Library trends 2: 522 (1954). (p. 522-523).
433. Reference 432, p. 525-529.
434. Lowry, K. W. Aeronautical documentation in the United States. Special libraries 47: 70-73 (1956). (p. 73).

Volume Four Part Five

CODING IN YES-NO FORM

by

Doralyn J. Hickey

Mechanical Representation of Codes in Yes-No Form

The simplest form of yes-no coding is achieved by allotting one specific "yes" position to represent one and only one concept (1). The polar character of the yes-no relationship implies that the total body of material to be coded can be divided into two distinct parts according to whether or not it is found to contain a given concept. For instance, books checked out of a library may be divisible into two main categories: reserve books and non-reserve books; then, one specific code position marked "yes" will indicate that a particular title belongs to the reserve section, while the same position (on another card) marked "no" will indicate that the second title does not belong to the reserve section (2). On the other hand, a concept which allows for significant variation in its components must code into more than one position. A typical example of this practice is the direct code designed for recording material on the geology and chemistry of coal; in this case the headings of major interest are coded directly on one section of the coding medium while subheadings are given additional positions on another section (3).

Although the authorities (4, 5) define direct coding in terms of its ability to separate materials into two groups--"a" and "not a"--this phrase is also used to designate systems in which the meaning of the code position is directly readable simply by examination of the code medium (6). Almost all of the examples of direct coding are to be found in connection with the use of the marginal punched card, and in the majority of cases the meaning of each position on the card is printed adjacent to the code position itself (7, 8, 9, 10, 11, 12, 13). In some cases, however, direct coding has been em-

ployed through the use of a printed number next to the code position, the meaning for the number being listed on a separate code index (14) or on a template (15).

The marginal punched card operates upon a relatively simple principle: small holes are perforated close to the edge of the card stock, and meanings are assigned to these holes; if the material being indexed is represented by one of the holes, then this position is notched out to the edge of the card. When a group of cards is sorted, the cards are superimposed and a small rod or needle inserted through the desired hole; as the needle is lifted, cards with a notch at that hole will fall off and can be easily separated from the rest (16, 17). Figure 1 shows an example of the marginal punched card employing direct coding. A number of different brands of cards are available, but the basic principle of operation is the same (18).

A modified form of direct coding may also be applied to cards designed for machine sorting (19, 20). In this case the design of the code bearer is responsible for the modification since the cards may be punched over their entire surface (21). These cards are produced in three basic patterns: the IBM card with 80 vertical columns and 12 positions per column; the Remington Rand card with 45 columns of 12 positions each (or 90 columns of 6 positions each) (22); and the Powers-Samas card with 40 columns of 12 positions each (23, 24). Examples of the three main types of cards are shown in Figures 2, 3, and 4.

It is not impossible to make use of straight direct coding on machine sorted cards; indeed, two writers record situations in which the positions on an IBM card were regarded as unique locations and a direct coding scheme was introduced (25, 26). This practice, however, often involves the punching of more than one position per column, which in turn complicates the machine sorting operation since some of the machines are not equipped to sort for more than one position per column (27, 28, 29). The alternative, of course, is to make use of a modified direct code in which the columns express given concepts and the positions in the columns

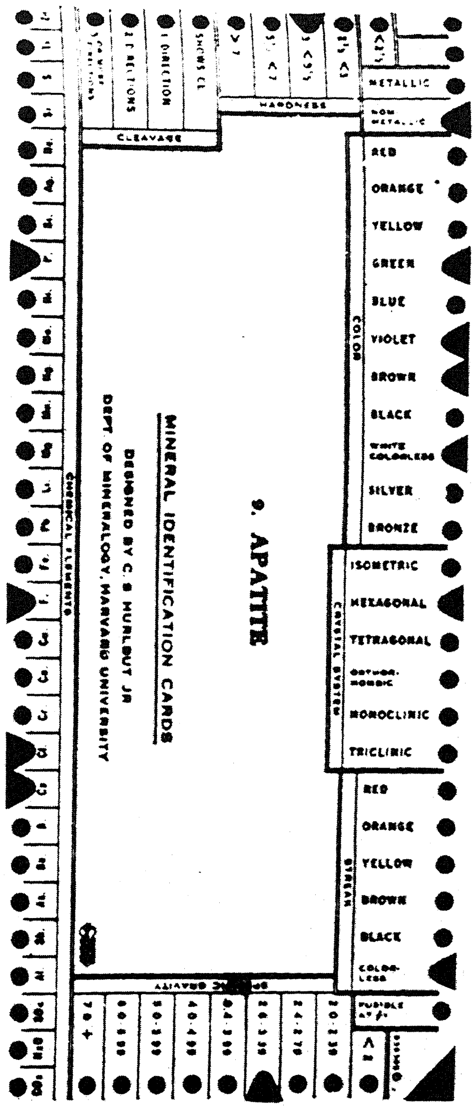


Figure 1

Direct Code on a Marginal Punch Card

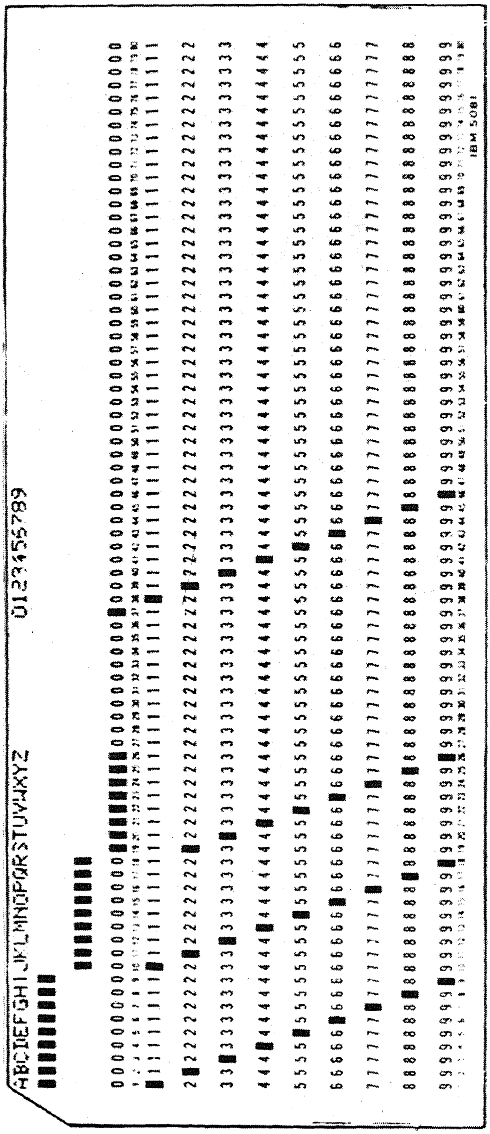
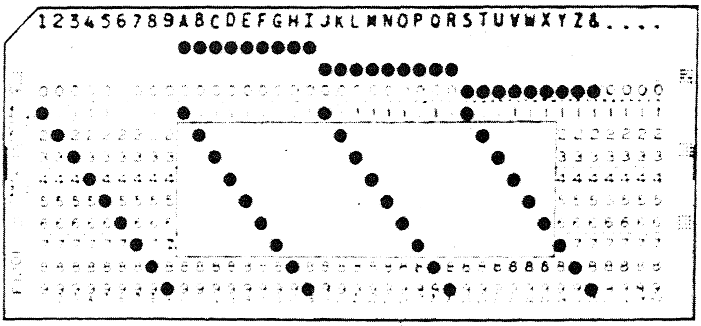


Figure 2

IBM Card, Actual Size, Indicating Punching Code



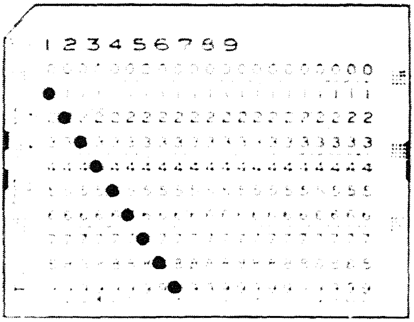


Figure 3

Underwood Samas Cards, Actual Size, Indicating Punching Code

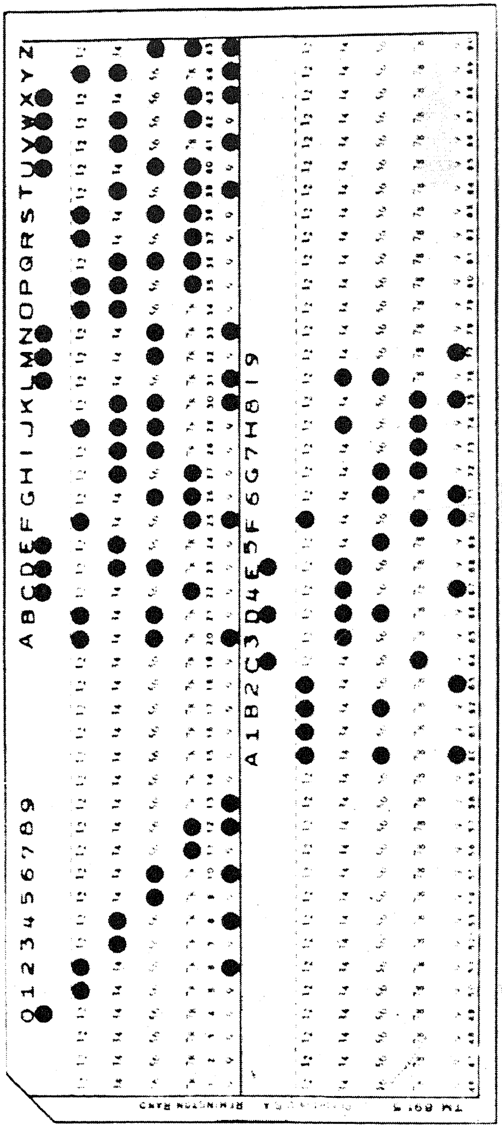


Figure 4

Remington Rand Card, Actual Size
Indicating Punching Code

are used to designate mutually exclusive subheadings (30). In the case that the subheadings are not necessarily mutually exclusive, the problem of double punching may arise again (31); however, solutions have been discovered to overcome this difficulty, e.g. rewiring the electrical circuits (32), inactivating certain contacts or feeding the cards in upside down (in the case of two punches per column) after they have been run in the normal way (33).

It is rare to find situations in which the machine sorted card contains direct coding with the complete meaning of the codes printed on the face of the card; however, the significance of the columns is sometimes printed directly onto the card (34, 35), and one writer describes a card which allows both the written and the coded form to appear on its face (36).

A code medium similar to both the machine sorted and the marginal punched cards is the slotted card (see Figure 5) which is designed for needle sorting, but which makes use of the entire face of the card rather than just the edge (37). The "yes" indication is achieved by slotting away the section of the card between two holes, the holes being numbered consecutively (38). Direct coding is applicable to this medium as well as to the marginal punched cards (39).

A fourth medium, which differs from those already mentioned by virtue of its form and of the use to which it is put, is the "Peek-a-boo" card, sometimes called the "Batten" card (40) and elsewhere known as the Cordonnier system (41). This system uses one card to represent each subject heading, and positions marked over the entire face of the card to stand for document numbers; if the subject heading pertains to a given document the pre-assigned number for that document will have its position punched out on the card (42). Documents which deal with each of several subjects may be located by superimposing the subject cards and observing the numbers of those positions which allow light to pass through (43). The number of documents which can be indexed per card and the size of the punch used, its placement on the card, etc., vary widely with the particular sys-

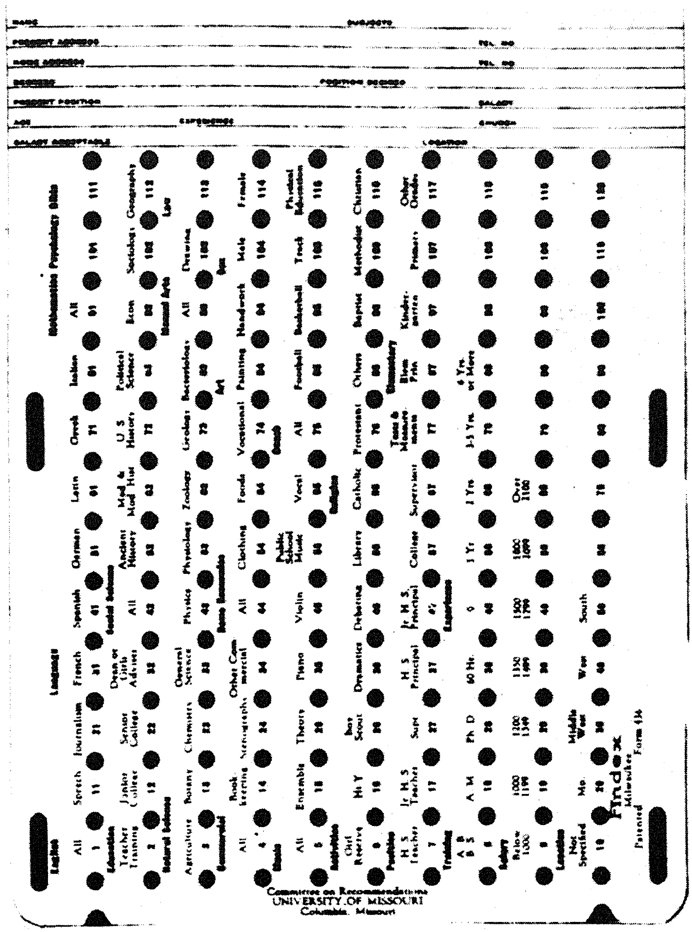


Figure 5

Direct Coding on a "Findex" (Slotted Card

tem chosen, but the coding is essentially direct in all cases (44, 45, 46).

Systems Using One or More Positions Per Concept Represented

Before summarizing the developments in the field of numerical and alphabetic coding, it may be helpful to outline the mathematical theory of combinations so that the coding capacity of the various media may be understood. No attempt will be made to prove the following statements; the reader is referred to items no. 47, 48, and 49 in the notes and also to item no. 1, p. 67-68 and no. 50, p. 276-283.

The basic formula for determining the number of possible combinations of "n" different items taken "m" at a time (where the order in which the items are placed has no effect upon the meaning of the aggregate) is expressed by the following equation:

$$(1)\quad C^{n}_{m} = \frac{n(n-1)(n-2)\ \dots\ (n-m+1)(n-m)(n-m-1)\ \dots\ (2)(1)}{[m(m-1)(m-2)\dots(2)(1)\ \ (n-m)(n-m-1)\ \dots\ (2)(1)]}$$

$$= \frac{n!}{m!(n-m)!} = \frac{n(n-1)(n-2)\ \dots\ (n-m+1)}{m!}$$

Provided that n and m are positive integers and n is greater than m.

Also:

$$(2)\quad C^{n}_{m} = C^{n}_{n-m} \qquad \text{and} \qquad (3)\quad C^{n}_{0} = C^{n}_{n} = 1$$

If a rearrangement of the order in which the items are taken changes the meaning of the aggregate, then the number of possible aggregates (or permutations) is determined by the following formula:

$$(4) \quad P^{n}_{m} = C^{n}_{m}(m!) = \frac{n!}{(n-m)!} = n(n-1)(n-2)\ \ldots\ (n-m+1)$$

In certain situations it is necessary to know how many different patterns can be obtained from "n" positions when each position can take on any one of "r" different integral values and the positions may be taken any number "m" (where "m" is an integer, $0 \leq m \leq n$) at a time. The formula becomes:

$$(5) \quad C = r^{n}$$

In the case that $r = 2$, then

$$(6) \quad C_{(2)} = \sum_{m=0}^{m=n} C^{n}_{m} = 2^{n}$$

Numerical Coding

A type of coding closely akin to direct coding is the representation of the ten decimal digits--0, 1, 2, 3, 4, 5, 6, 7, 8, 9--by means of ten positions on the coding medium; from several of these groups taken together, the ordinary decimal numbers may be built up, letting one group stand for the units position, another for the tens, a third for the hundreds, and so on (51). Although this procedure is applicable to marginal punched cards (52), the machine sorted cards use it more often and in a slightly different form: columns are used to represent the units, tens, hundreds, etc., positions in the decimal number, and the value of the digit is indicated by punching the desired row in the proper column (53). In some cases the numbers punched into the cards in this manner are the actual quantities which are being recorded (54); in other cases the numbers are themselves a code standing for certain subject material (55).

A related means for representing the ten decimal digits has been employed by telephone systems: each digit is transformed into a certain number of holes punched into a tape, the number of holes being equiva-

lent to the value of the digit (56).

A very common way to represent decimal digits through the use of less than ten positions is the "weighted" code, which has "the property that values, or weights, can be assigned to each of the ... bits [positions] with the decimal digit being represented equal to the sum of the weights" (57). Many forms of this code are available (58), but the 7-4-2-1 code is by far the most popular (59, 60, 61, 62, 63, 64). By means of this code, the numbers 1 through 14 may be represented, the number 6 being coded as a punch in the "4" position and a punch in the "2" position, for example (65). This code further has the property of facilitating sequential order, since a group of marginal punched cards may be sorted into numerical order simply by passing the needle through each of the four holes in sequence, each time placing at the end of the pack (in order) the cards which drop out (66, 67). Decimal numbers larger than 9 may be coded in either one of two ways: a second field of four positions (representing 70, 40, 20, and 10) may be added (with more fields if necessary, as the size of the decimal number warrants) (68), or a greater coding capacity may be secured by allowing each field to represent the full 14 numbers so that the "1" position in the second field would represent the number 15, etc. (69).

The application of the 7-4-2-1 code to machine sorted cards is relatively infrequent, although it has been suggested that 8 of the 12 positions per column could be used with this code for the representation of Dysonian chemical notation (70). A suggestion by Dunn (71) (which will be mentioned again later) has been transformed into the 7-4-2-1 pattern for machine sorted cards by Reagh (72), who recommends the 7-4-2-1 code as simpler, although it reduces the coding capacity of a column from Dunn's figure to 1, 500.

The 7-4-2-1 combination is by no means the only available 4-place code; Richards (73) lists 17 weighted 4-bit (i.e., 4-position) codes, some of which have the additional property of being "self-complementing" (74). This latter factor is particularly useful with regard to

digital calculating machines since it means that the 9's complement (the decimal number which, when added digit by digit to a given number will yield a total of 9 in each decimal position) of a digit is obtained by changing the state of each of the switches used to represent the digit. As Richards puts it, "The 9's complement of each decimal digit may be obtained by changing the 1's to 0's and the 0's to 1's in the coded representation of the digit" (75).

A 4-position code which behaves in almost the same manner as the more conventional 7-4-2-1 code is the 8-4-2-1 system (76). This is actually a 4-place development of the straight binary representation (77) or geometrical progression (78). It is not, however, a self-complementing code, although it can be transformed into one by the following process (79): "In this [excess-3] code, 3 is added to each decimal digit to give an excess-3 value which is then represented by a corresponding four-place binary number" (80). Hartree cites Stibitz as the originator of the excess-3 code and indicates that it has three advantages over straight binary representation:

It gives a positive indication of the digit zero, complements are obtained by interchanging 01s and 1's, and in the process of addition carry-over from the most significant of the four binary digits occurs just when, and only when, carry-over from the corresponding decimal place is required (81).

The 8-4-2-1 code has also been recommended as a means for increasing the capacity of the columns on machine sorted cards; Royer (82) suggests the division of the 12 positions in each column into three groups of 4 each, assigning the 8-4-2-1 code to each group. Eckert (83) records the use of the 8-4-2-1 code in the IBM Selective Sequence Electronic Calculator.

Codes using more than four positions for the representation of a single decimal digit have been developed to simplify certain special operations involving marginal punched cards (84) and computing machinery (85). A common scheme which facilitates the selection

of individual marginal punched cards without sacrificing the easy ordering feature is the SF-7-4-2-1-0 system (86). The "SF" position indicates that the other punch in the field stands for a single figure; this position is not punched when the decimal digit represented is 3, 5, 6, 8, or 9 (87). The use of the "0" in the code becomes essential only when more than one field is employed; its function is to separate, for example, the 70's from the 71's, 72's, 73's, etc. (88). The essential characteristic of a selector code, however, is that the same number of yes-designations be placed in the field for each digit represented; and since five positions taken two at a time yield exactly ten distinct meanings, the 5-bit selector code is widely used (89, 90). In the case of the marginal punched card, a 5-position selector code provides for the isolation of an individual digit by two simultaneous needlings (91). Of course many other selector codes are possible; a table of the various combinations and their capacities is included by Wise (92).

Selector codes on marginal punched cards are sometimes equipped for direct reading by the use of a triangular arrangement; the basic principle of the code, however, remains the same (93). For a diagram of triangular selector coding, see Figure 6. The meaning of

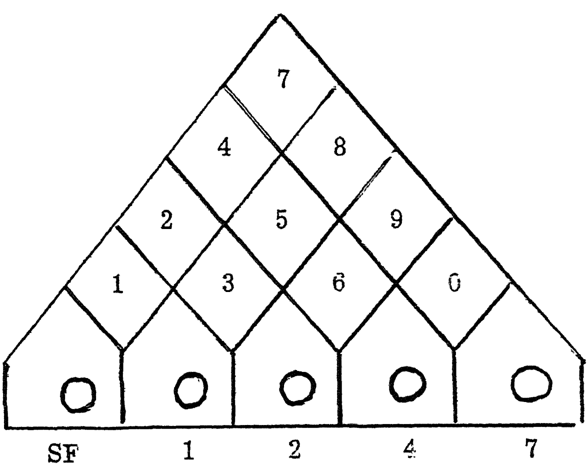


Figure 6

5-Position Triangular Code for a Marginal Punch Card

any two punches may be determined by reading the number printed in the triangle as the intersection of the diagonal columns marked off by the punches.

Five-bit selector codes also offer advantages for use in electronic calculators: "With 5 bits there are ten different combinations with two 1's, and when the ten decimal digits are represented in this manner, it is possible to distinguish any digit without the use of inverters, which is an advantage" (94). The IBM Magnetic Drum Calculator type 650 employs, is one part of its operation, a selector code of the form 6-3-2-1-0 (95), although developers of one of the Harvard calculators have found a 5-bit weighted 8-6-4-2-1 code to be useful (96). The Rapid Selector, which records its code by means of opaque and transparent spots on film, employs a 5-position selector code with an overlapping arrangement which will record decimal numbers to a magnitude of seven places (97). A diagram of the code is shown in Figure 7.

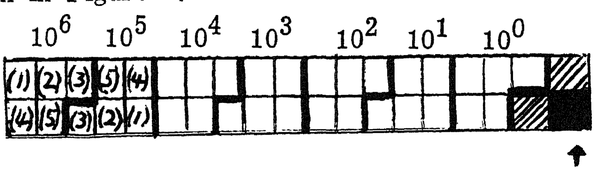


Figure 7

Synchronization mark

Arrangement of the 7-field 5-position Rapid Selector Code. Heavy lines indicate the division of the fields; numbers in parenthesis show the location of the positions in the two types of fields. The synchronization mark positions the rows properly for scanning.

The use of codes of five or more bits to represent decimal digits for purposes of machine calculations has been suggested for two reasons: "[the use of them may make it] possible to effect simplifications in the arithmetic circuits in some cases ... [and they can provide] the ability to detect errors" (98, 99). Krider reports the use of a 6-position code for automatic programming of certain small calculators (100), and Richards goes into some detail about error-detecting and error-correcting codes which may use seven or eight bits to repre-

sent the ten decimal digits (101). Such error-detecting and correcting codes are subject to the following conditions (102): a single-error-detecting code requires the use of five bits; single-error-correcting, seven bits; double-error-detecting, eight bits. Another way of stating the conditions is that the detecting and correcting powers increase with the number of changes required to transform the representation of one digit into the representation of another; if only one change is required, then no error-detecting or error-correcting properties are available. A code having the desired detecting and correcting qualities may be constructed through the use of "redundancy bits": extra positions used to indicate whether or not the quantity of 1's in a certain section of the coded digit is even or odd; if there are an even number of 1's, then the redundancy bit is a 1, but if there are an odd number of 1's, the redundancy bit is a 0 (103).

Up to this point we have been discussing the various means by which decimal digits (0 through 9) may be converted into code patterns using only yes-no designations. The fact that "almost all high-speed computers are based on two-state storage systems" (104) has suggested that "A simpler method of using two-state elements is to work in the scale of two, or the 'binary' scale, as it is often called" (105). Livesley gives the following comparative table of numerical values in the decimal and binary forms:

<u>Decimal Form of Number</u>

6 $(= 6 \times 10^{0})$
18 $(= 1 \times 10^{1} + 8 \times 10^{0})$
273 $(= 2 \times 10^{2} + 7 \times 10^{1} + 3 \times 10^{0})$
3.25 $(= 3 \times 10^{0} + 2 \times 10^{-1} + 5 \times 10^{-2})$

<u>Binary Form</u>

110 $(= 1 \times 2^{2} + 1 \times 2^{1} + 0 \times 2^{0})$
10010 $(= 1 \times 2^{4} + 1 \times 2^{1})$
100010001 $(= 1 \times 2^{8} + 1 \times 2^{4} + 1 \times 2^{0})$
11.01 $(= 1 \times 2^{1} + 1 \times 2^{0} + 1 \times 2^{-2})$

Richards points out that the use of both binary and decimal systems for computing machinery is very common:

> So far as is known, radices two, three, eight, twelve, and sixteen are the only ones which have ever received serious consideration for use in computing machinery. The list of those which have actually been used is much more restricted. In fact, no computer is known in which a radix other than two or ten is employed. One minor exception to this last statement exists in that at least two companies have built small electromechanical desk computers in which radix eight is used (107).

He points out, however, that "printed numbers in the binary system are undesirable because it is difficult to handle a large number of nothing but 0's and 1's without making excessive errors" (108).

Upon occasion, a binary system has been recommended for use with machine sorted cards; thus, the twelve positions in each column could be used in binary fashion to represent decimal numbers from 1 to 4,095 (109). This procedure, however, reintroduces the problems of multiple punches per column which were discussed earlier. The use of straight binary, and even ternary, coding on marginal punched cards has been suggested by Giffler (110), but there is no evidence that such schemes have been put into common practice.

There are two other radix systems which have certain features to commend them. The octonary system has been suggested for the following reasons:

> The expression of a number in binary notation requires a comparatively large number of digits .... For example, a number with six decimal digits may require 20 binary digits; the same number, in octal notation, may need only seven. The number of octal digits is thus only slightly greater than the number of decimal digits, the conversion from binary representation

is trivially easy [see table below], and an inexperienced reader can get good qualitative estimates of the results without worrying about the octal character of the representation (111).

Binary Representation of Octal Digits

Binary triplet	Octal mark
000	0
001	1
010	2
011	3
100	4
101	5
110	6
111	7

A system of mixed radices--the biquinary system --has been employed by the Bell Telephone Laboratories in some of their computers (see reference no. 80, p. 76). Richards asserts: "The biquinary code has 7 bits with the weights of 5, 0, 4, 3, 2, 1, 0. With this code, arithmetic operations may be performed in a moderately straightforward manner, although whether or not there is a net simplification when compared with the 4-bit codes is a debatable point. The main reason for the use of 7 bits is the ability to detect errors" (112). Hamilton records the use of this code in one portion of the IBM Magnetic Drum Calculator (type 650) (113). Richards also mentions a quibinary code with weights 8, 6, 4, 2, 0, 1, 0, but he does not cite any computer in which this code has been used (114).

The desire to gain coding capacity on marginal punched cards has led to experiments with systems using more than one row of holes (115). McGaw reports the use of a 4-position, 2-row field which provides for "the selective sorting of nine classifications" (116). Ruston discusses the same system in terms of triangular coding, with each intersection representing two digits: if the upper digit is to be indicated, the left hand column is deep-punched and the right hand column is shallow-punched, while the reverse arrangement indicates the lower digit (see Figure 8) (117). Draheim goes into

great detail about the various coding possibilities of ten positions in double and triple row combinations (118); however, the use of double and triple rows has not been standardized to any great degree as yet (119). One fairly common use seems to be the assignment of the deep punches to a broad subject and of the shallow punches to subheadings (120). The 3-row system may also employ a third type of punch--much like the slotting process described earlier--to increase the capacity and selectivity features of a marginal punched card (121), although the sorting procedure may be considerably complicated thereby (122).

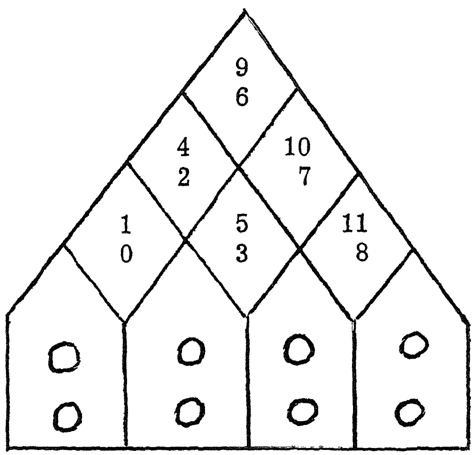


Figure 8

Triangular 2-row, 4-position Coding
on a Marginal Punch Card

Coding the Alphabet

The representations of alphabetic material in yes-no form have been designed with two main objects in mind: the coding of individual letters as separate entities and the coding of letters so as to form words (123, 124). Although marginal punched cards have been designed which simply allot 26 spaces for the letters of

the alphabet (125), other more economical systems have been developed. Among these is one which has received considerable attention: the OLECB plan developed by Gerald Cox (126, 127, 128). This code has been issued in two versions, the revised plan appearing below (129, 130):

Letter	Code	Letter	Code	Letter	Code
A	No punch	M*	IE	U	OI
B	B	MAC	IEB	V	OIB
C	C	M**	IEC	W	OIC
D	CB	N	IECB	X	OICB
E	E	O	O	Y	OIE
F	EB	P	OB	Z	OIEB
G	EC	Q	OC		
H	ECB	R	OCB		
I	I	S*	OE		
J	IB	SCH	OEB		
K	IC	S**	OEC		
L	ICB	T	OECB		

*Used for the initial letter of words which alphabetically precede the next group of letters in the system.

**Used for the initial letter of words which alphabetically follow the preceding group of letters in the system.

The extra letter combinations have been included so as to facilitate the alphabetical arrangement of index words, and sequence sorting proceeds along the same lines as that for the 7-4-2-1 numerical code (131).

Another means of alphabetical arrangement actually uses the 7-4-2-1 system, plus an additional position marked "M" (132); the code is shown below:

Letter		Code
A	M*	1
B	N	2
C	O	1 & 2
D	P	4
E	Q	4 & 1
F	R	4 & 2

*All letters in the second column also receive a punch in a position marked "M".

Letter		Code
G	S	4 & 2 & 1
H	T	7
I	U	7 & 1
J	V	7 & 2
K	W	7 & 4
L	X, Y, Z	7 & 2 & 1

A third arrangement, which does not require the grouping of x, y, z, is the NZ-7-4-2-1 code; this code makes use of all combinations of 7-4-2-1 which total numbers 1 through 13, and these combinations are in turn allotted to letters A through M and to N through Z (with the additional NZ punch being used for the second group (133). The Dysonian system for representing letters by combinations of 1, 2, 4, 7, 10, 20, 40, 70, and a position marked L, on a single IBM card column, is very similar to the NZ-7-4-2-1 system, except that Dyson assigns 26 rather than 13 numbers to the letters and uses the "L" punch to distinguish letters from numbers (134).

A system which does not allow for sequence sorting is the 15-hole code reported by Anderson (135); this code allots one hole to each vowel and one hole to every two consonants, with x, y, z, again being grouped together. Westbrook and DeWald also suggest the assignment of two letters per hole--for the purpose of coding the initial letters of author's names (136).

A 9-position alphabetic selector code, made up of letters A, B, D, F, G, K, P, S, and V, has been reported by Wise as a system for representing the initial letter of an author's name. The code is as follows (137):

Letter	Holes Punched	Letter	Holes Punched
A	ASF	N	KD
B	BSF	O	KG
C	BA	P	PSF
D	DSF	Q	PA
E	DA	R	PB

Letter	Holes Punched	Letter	Holes Punched
F	DB	S	PD
G	GSF	T	PG
H	GA	U	PK
I	GB	V	VSF
J	GD	W	VA
K	KSF	X	VB
L	KA	Y	VD
M	KB	Z	VG

The use of double and triple rows on marginal punched cards for alphabetical coding has been suggested or reported by certain writers. Draheim describes a triple-row, 9-position arrangement, incorporating "sch" as his twenty-seventh letter (138). Campbell suggests a double-row, 3-position code for representing the first letters of authors' last names, observing that in his field of general organic research a sorting of the cards by author is seldom done, and thus some of the less commonly used letters may be combined into a single representation (139).

The triangular alphabetic code is discussed by Ruston (140), Hood (141), and Cox (142). The principles of this type of code are exactly the same as those for numerical triangular coding; a 2-row, 6-position version of the code is shown in Figure 9 (143).

Alphabetic representation on machine sorted cards has been, to a large degree, standardized, although each manufacturer presents a slightly different version (144). The form taken by these codes is indicated in Figures 2, 3, and 4.

IBM and Samas use a constant number of punches --two each--for the letters, while Rand uses both two and three punch combinations (145). Cochran (146), Gull (147), and Nalbandjan (148) all mention the use of the IBM alphabetic system; but Moffit is concerned about the amount of space required to code words in this manner (149), since only one letter can be coded into and interpreted from each column on the machine sorted card.

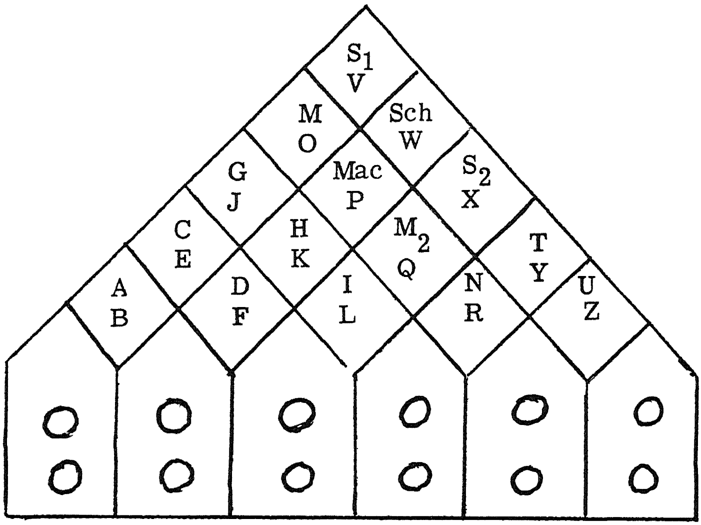


Figure 9

Triangular 2-row, 6-position Alphabetic Code on a Marginal Punch Card (Subscripts indicate the division of a letter into two coding parts to facilitate an alphabetic ordering.)

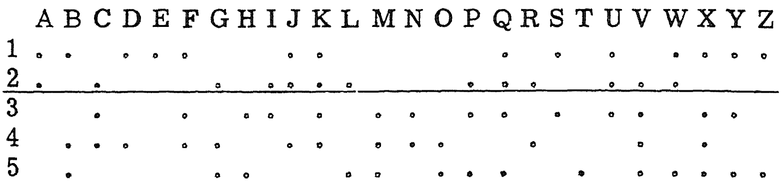
A different code system designed for use with machine sorted cards was suggested by Samain (150), who zoned the card into 6-column groups (omitting the first column). Each group thus contained twelve rows of six positions each. By allowing three perforations to each 6-position row, he could code any one of 20 different combinations into a single row in one group. Assuming that the alphabet could be reduced to 20 letters, the capacity of an 80-column, 12-row card would become 24 6-letter ideas.

The development of telegraphy has occasioned another type of alphabetic code, commonly known as Morse code (151). Although the code is essentially binary, the so-called "dash" being merely an extended "dot," the signal duration can be varied in different

ways, making the system ternary, quinary, or what have you, in its behavior (152). The use of time as a variable in the representation of decimal digits by means of binary equipment is also noted by Richards in connection with computing machinery (153). The American and International versions of the Morse code are shown below, together with the 2-channel tape version used for cable transmission.

	American	International: Land lines	International: Cable
A	.—	.—	$\begin{array}{cc}\cdot & \\ \hline & \cdot\end{array}$
B	—...	—...	$\begin{array}{cccc} & \cdot & \cdot & \cdot \\ \hline \cdot & & & \end{array}$
C	..—.	—.—.	$\begin{array}{cccc} & \cdot & & \cdot \\ \hline \cdot & & \cdot & \end{array}$
D	—..	—..	$\begin{array}{ccc} & \cdot & \cdot \\ \hline \cdot & & \end{array}$
E	.	.	$\begin{array}{c}\cdot \\ \hline \ \end{array}$
F	.—.	..—.	$\begin{array}{cccc}\cdot & \cdot & & \cdot \\ \hline & & \cdot & \end{array}$
G	——.	——.	$\begin{array}{ccc} & & \cdot \\ \hline \cdot & \cdot & \end{array}$
H	....	....	$\begin{array}{cccc}\cdot & \cdot & \cdot & \cdot \\ \hline & & & \end{array}$
I	..	$\begin{array}{cc}\cdot & \cdot \\ \hline & \end{array}$	..
J	—.—.	.———	$\begin{array}{cccc}\cdot & & & \\ \hline & \cdot & \cdot & \cdot\end{array}$
K	—.—	—.—	$\begin{array}{ccc} & \cdot & \\ \hline \cdot & & \cdot\end{array}$
L	——(5)	.—..	$\begin{array}{cccc}\cdot & & \cdot & \cdot \\ \hline & \cdot & & \end{array}$
M	——	——	$\begin{array}{cc} & \\ \hline \cdot & \cdot\end{array}$
N	—.	—.	$\begin{array}{cc} & \cdot \\ \hline \cdot & \end{array}$
O	. .	———	$\begin{array}{ccc} & & \\ \hline \cdot & \cdot & \cdot\end{array}$
P	.....	.——.	$\begin{array}{cccc}\cdot & & & \cdot \\ \hline & \cdot & \cdot & \end{array}$
Q	..—.	——.—	$\begin{array}{cccc} & & \cdot & \\ \hline \cdot & \cdot & & \cdot\end{array}$
R	. ..	.—.	$\begin{array}{ccc}\cdot & & \cdot \\ \hline & \cdot & \end{array}$
S	...	...	$\begin{array}{ccc}\cdot & \cdot & \cdot \\ \hline & & \end{array}$
T	— (3)	—	$\begin{array}{c} \ \\ \hline \cdot\end{array}$
U	..—	..—	$\begin{array}{ccc}\cdot & \cdot & \\ \hline & & \cdot\end{array}$
V	...—	...—	$\begin{array}{cccc}\cdot & \cdot & \cdot & \\ \hline & & & \cdot\end{array}$
W	.——	.——	$\begin{array}{ccc}\cdot & & \\ \hline & \cdot & \cdot\end{array}$
X	.—..	—..—	$\begin{array}{cccc} & \cdot & \cdot & \\ \hline \cdot & & & \cdot\end{array}$
Y	.. ..	—.——	$\begin{array}{cccc} & \cdot & & \\ \hline \cdot & & \cdot & \cdot\end{array}$
Z	... .	——..	$\begin{array}{cccc} & & \cdot & \cdot \\ \hline \cdot & \cdot & & \end{array}$

The 5-channel punched tape code used by the teletype is sometimes called the Baudot code and is as follows:



(Line indicates position of feed holes)

Some of the common language machines have been designed to operate from 5-channel tape because of the equipment already in production for translating alphabetical and numerical data into that form, even though 6, 7, and 8-channel tapes are also available (154).

An alphabetic code based on the binary system has been adopted as a standard means of producing books for the blind (155). The Braille alphabet is composed of groups of six embossed dots arranged in two columns of three dots each (see Figure 10). In advanced Braille, the dots may be used to indicate combinations of letters rather than individual ones, but the system is essentially the same (156). The number of combinations available in this system is 63 (157, 158).

Because of the time required by the operation of converting written or typed data into a form suitable for computer input, means have been developed to by-pass this extra step under certain limited conditions. A photoelectric scanning device is now capable of transforming printed characters directly into binary form since each character is made up of a distinctive pattern or horizontal and vertical marks. Whenever the scanner detects a black mark on the white background, it produces an electrical impulse; and the complete impulse pattern becomes the binary representation of the character (159, 160). The reverse of this procedure may also be accomplished: binary machine code may be reconverted into alphabetic symbols through the use of a matrix of wires. Bursts of current are sent through the matrix at various points which in turn deposit charges on a paper in contact with the matrix; ink is attracted to the charged points on the paper, producing the form of the alphabetic character (161). Another sys-

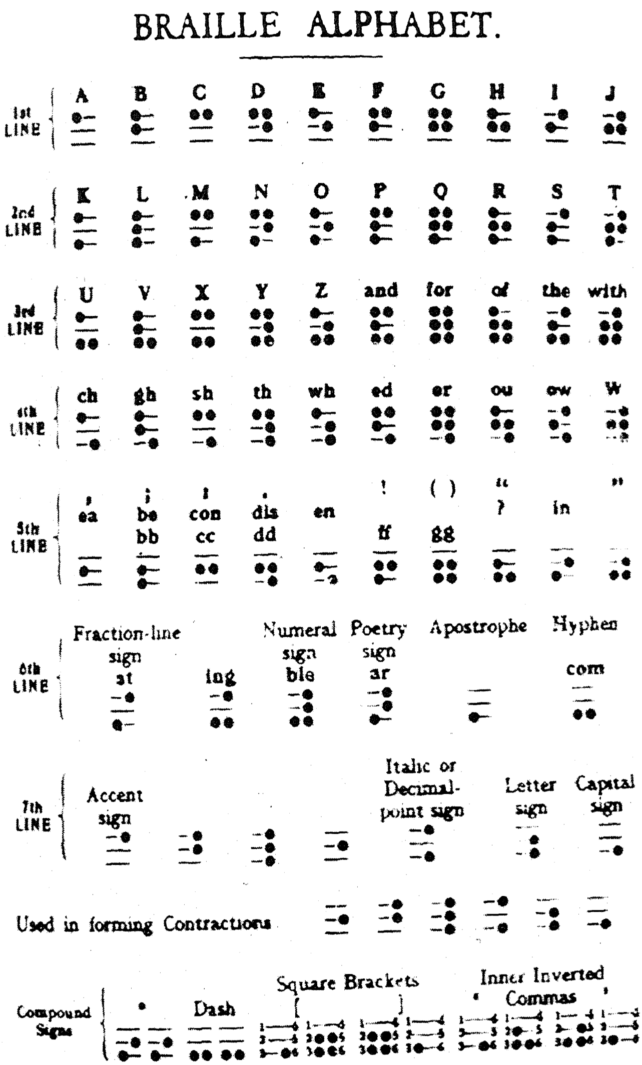

BRAILLE ALPHABET.
1st LINE
A B C D E F G H I J
2nd LINE
K L M N O P Q R S T
3rd LINE
U V X Y Z and for of the with
4th LINE
ch gh sh th wh ed er ou ow W
5th LINE
, ea
; be bb
: con cc
. dis dd
en
! ff
( ) gg
" ?
in
"
6th LINE
Fraction-line sign st
ing
Numeral sign ble
Poetry sign ar
Apostrophe
Hyphen com
7th LINE
Accent sign
Italic or Decimal-point sign
Letter sign
Capital sign
Used in forming Contractions
Compound Signs
Dash
Square Brackets
Inner Inverted Commas

Figure 10

tem converts electrical impulses into readable characters on a telescreen (162).

Superimposed Coding

The numeric and alphabetic codes discussed thus far do not, in some situations, produce enough possible combinations to cover all of the categories which may need to be coded. To solve this problem, a type of coding called "superimposed" has been developed (163). Two major forms of such coding have been suggested: an alphabetic system called "word coding" (164) and a random number system called "Zatocoding" (165). The basic principle of both systems is, however, the same, for they each use codes made up of a given number (or numbers) or characters and punch (or otherwise record) the characters representing several different items into the same field. When this is done, overpunching may occur since the codes for any two items can have certain characters in common. Once a character (i. e., a position) is punched, a second punch cannot be detected; hence, the total number of punches in the field may be less than the sum of the characters used to represent the items (166, 167).

The overlapping of codes can result in an inability to make an absolute selection of cards which deal only with the item wanted; in other words, the patterns of several items may overlap enough to produce the pattern of another item with which the card has no connection. When the cards are sorted for this latter item, so-called "extra cards" may appear (168). Because an excessive number of unwanted cards would impair the value of the system, both Wise and Mooers have worked out mathematical theories which are intended to keep the unwanted cards to a minimum (169, 170). For various discussions of the mathematical theory involved in these systems, the reader is referred to the following sources: 50, 163, 169, 171.

The conclusions reached by Mooers and by Wise may be summarized as follows:

Mooers--"The sum of the separate punches of the individual code patterns placed in the coding field of a card shall not exceed 69% of the total number of positions in the field, in which case the average density of punched positions in the field will not exceed 50%. This is the condition indicating optimum utilization.... With a selection according to S positions, the ratio of the number of extra cards to the total number of cards sorted will be less than $(1/2)^S$ on the average" (172).

Wise (with reference to word coding)--"We may say that for practical punching schemes involving a maximum of entries punched on the card the optimum conditions are attained when the number of instructions to punch in a given field amounts to approximately 46 per cent of the available punching positions. Because of overlapping, such a proportion of punching instructions will result in approximately 37 per cent ($1/e$) of the positions being actually punched" (173).

Wise (with reference to random number codes, where X is the total number of holes punched and H is the total number of holes available on the medium)--"It should be noted that the minimum optimum value of X/H is 0.500000 or (1/2) instead of 0.367879 ($1/e$).... Both systems of punching have the value of X/H approaching a maximum optimum of 0.693147 as the number of subjects increases" (174).

The 69% appears as a common value determined by both Wise and Mooers. A further selection from Mooers' writings will serve to show a practical application of his conclusions:

According to the method of Zatocoding we will first compute the vocabulary V, and then the number of combinations T. In a field of F positions, in which N punches per code pattern are placed without restriction upon their position, the number of different patterns, and thus the possible coding vocabulary, is ${}_F C_N$ or the number of combinations of F things taken N at a time. Thus the coding vocabulary V in a 40 position field with a four-punch code is 91,390 different codes. Other code lengths may of course be used. The capacity of this

field of 40 positions is obtained by finding 69% of 40, which is 27.6, or nearly 28. Therefore, the field can contain seven four-punch codes (because four times seven equals twenty-eight). By the Zatocoding statistics we know that on the average 20 positions on the field will be punched, due to expected overlapping, when these 28 code punches are placed on the field. Seven codes out of a vocabulary of 91,390 different ones may be placed in the field. Then the number of combinations T possible on the field of a card is the number of combinations of 91,390 things taken seven at a time, or something in the order of $(10)^{31}$ (175).

A certain amount of disputation has developed between Mooers and Wise upon the relative merits of their respective systems (176, 177, 178). Mooers rejects the alphabetical system by Wise because it lacks randomness (179) and further charges that Wise's application of superimposed coding to the Rapid Selector is mathematically in error (180, 181). In a later comment, Wise suggests that the alphabet might be given a random distribution by increasing it to 30 positions and applying the author analysis used by Cox, Casey, and Bailey in setting up their alphabetical code (182). With regard to the Rapid Selector, Wise would use a 6-letter code and up to 16 subjects per item, with a search directed to a single code yielding extra cards to a ratio of less than 1 to 100; a search on two codes, less than 1 to 10,000; a search on three codes, less than 1 to 1,000,000 (183). Under Mooers' calculations, the most efficient code would use a random pattern of 8 positions, with 18 patterns (or subjects) allowable per field; a selection on the basis of one pattern would yield extra cards in a ratio of less than 1 to 500; on two patterns, less than 1 to 50,000; on three patterns, less than 1 to 5,000,000 (184).

Two processes for selecting random numbers suitable for use in superimposed codes have been described (185, 186); one of which makes use of the Rand Corporation's publication of 1,000,000 random digits (187). The Brown and Oneal paper applies the system to IBM cards, rather than to the marginal punched cards used by Mooers (189) and Wise (190).

Both Mooers (191) and Wise (192) claim that superimposed systems are highly successful when constructed according to the mathematical specifications. Several installations based on the principles set forth by Mooers, some of which also use IBM rather than marginal punched cards, are reportedly functioning properly and satisfactorily (193, 194, 195). Support for the word coding system is furnished by articles from Sebring and from Perry (196, 197).

References

1. McGaw, Howard F. Marginal punched cards in college and research libraries; Washington, D. C., Scarecrow press, 1952, p. 44.
2. Reference 1, p. 45.
3. Breger, Irving A. Applications of simple coding procedures to a specific problem. In: Casey, Robert S. and James W. Perry, Punched cards; New York, Reinhold, 1951, p. 30-35.
4. Reference 1, p. 44.
5. Casey, Robert S., C. F. Bailey, and Gerald J. Cox. Punch card techniques and applications. Journal of chemical education, 23:495-99 (1946).
6. Reference 3, p. 34.
7. Clarke, S. H. A multiple-entry perforated-card key with special reference to the identification of hardwoods. New phytologist, 37:369-74 (1938).
8. Dadswell, H. E. and Audrey M. Eckersley. The card sorting method applied to the identification of the commercial timbers of the genus Eucalyptus. Council for scientific and industrial research (Australia), Journal, 14 (no. 4):266-80 (1941).
9. Dunkley, H. L. A multiple-entry perforated card-key for the identification of Uganda trees. Empire forestry journal, 18:83-90 (1939).
10. Epstein, Albert. Statistical analysis with hand punched and sorted cards. American statistician, 1 (no. 2):6-7 (Oct. 1947).
11. Hurlbut, C. S., Jr. Mineral determination by means of punched cards. American mineralogist, 33:508-12 (1948).
12. Making alloy information accessible. Metal industry (London), 54:617-8 (Jan. -Je. 1939).
13. Normand, D. Les clés pour l'identification des bois et le systeme des fiches perforées. L'agronomie tropicale, 1:162-72 (1946).

14. Orton, Floyd E. Let's look at paper work. Library journal, 75:368, 384-87 (1950).
15. Diakonoff, A. Het paramount-kaartsysteem ten dienste van de entomologie. Entomologische mededeelingen van Nederlandsch-Indië, 7 (no. 2): 34-40 (1941).
16. Reference 1, p. 15.
17. Casey, Robert S. and James W. Perry. Introduction, in their Punched cards; New York, Reinhold, 1951, p. 5.
18. Casey, Robert S. and James W. Perry. Elementary manipulations of hand-sorted punched cards, in their Punched cards; New York, Reinhold, 1951, p. 10.
19. Black-Schaffer, Bernard and Paul D. Rosahn. Methods of analysis of Yale autopsy protocols, including a code for the punched card study of syphilis. Yale journal of biology and medicine, 15:575-86 (1942/43).
20. Eckert, Wallace John. Punched card methods in scientific computation; New York, Thomas J. Watson astronomical computing bureau, Columbia University, 1940.
21. Reference 17, p. 5.
22. Lawson, Murray G. The machine age in historical research. American archivist, 11:141-49 (1948).
23. Friedman, Burton Dean. Punched card primer; Public administration service, 1955, p. 8-13.
24. Powers-Samas cards are also available, according to Friedman (see reference 23) in the 21 column size, while Callander (Punched card systems; their application to library technique. Library association record, 48:171-4((1946))) describes a card, issued by the same company, which has 36 columns of 10 positions each.
25. Gage, N. L. and H. M. Remmers. Opinion polling with mark-sensed punch cards. Journal of applied psychology, 32:88-91 (1948).
26. Taylor, F. Lowell. Numerical index key for the Beilstein system. Industrial and engineering chemistry, 40:470-73 (1948).
27. Edwards, Thomas I. The coding and tabulation of medical and research data for statistical analysis. Public health report (U.S. Public Health Service)

57 (pt. 1):7-20 (Jan. -Je. 1942).
28. Kempthorne, O. The use of a punched-card system for the analysis of survey data, with special reference to the analysis of the National farm survey. Royal Statistical Society, Journal, 109: 284-95 (1946).
29. Peakes, Gilbert L. Report indexing by punched cards.. Journal of chemical education, 26:139-46 (1949).
30. Reference 19.
31. Roches, Camille. La mécanographie au service de la documentation. L'industrie textile, 63:107-8 (1946).
32. Appel, Valentine and George Cooper. A refinement in the use of mark-sense cards for test research. American statistical association, Journal, 50:557-60 (Je. 1955).
33. Peakes, Gilbert L. Report indexing by machine-sorted punched cards. In: Casey, Robert S. and James W. Perry, Punched cards; New York, Reinhold, 1951, p. 119.
34. Fletcher, Mona. The use of mechanical equipment in legislative research. American Academy of Political and Social Science, Annals, 195:168-75 (1938).
35. U. S. Department of Commerce, Bureau of the Census, United States census of agriculture, 1935, Descriptive supplement: technique of tabulation; Washington, D. C., U. S. Government printing office, 1937.
36. Berkson, Joseph. A punch card designed to contain written data and coding. American statistical association, Journal, 36:535-38 (1941).
37. Reference 1, p. 198-201.
38. Reference 1, p. 200.
39. National Research Council, Research information service, Mechanical aids for the classification of American investigators, with illustrations in the field of psychology, by Harold C. Bingham; Washington, D. C., National Research Council, 1922.
40. Batten, W. E. Specialized files for patent searching. In: Casey, Robert S. and James W. Perry, Punched cards; New York, Reinhold, 1951, p. 171.

41. Schürmeyer, Walther. Selecto - ein neues auswahlsystem für die dokumentation. Nachrichten für dokumentation, 3:33-34 (1952).
42. Batten, W. E. A punched-card system of indexing to meet special requirements. Association of special libraries and information bureaux, Re- of the proceedings, 22:37-39 (1947).
43. Reference 42.
44. Reference 41.
45. Reference 42.
46. Loosjes, Th. P. The Delta card. Aslib proceedings, 9 (no. 5):142-43 (May 1957).
47. Whitworth, William Allen. Choice and chance; New York, Stechert, 1934, p. 1-67.
48. Burington, Richard Stevens and Donald Curtis May, Jr. Handbook of probability and statistics with tables; Sandusky, O., Handbook publishers, p. 25-27.
49. Briggs, William and G. H. Bryan. The tutorial algebra, rewritten by George Walker; London, University tutorial press, 1954, 1: 327-370.
50. Wise, Carl S. Mathematical analysis of coding systems. In: Casey, Robert S. and James W. Perry, Punched cards; New York, Reinhold, 1951.
51. Draheim, Heinz and Ottokar Gdaniec. Was leisten zehn löcher einer randlochkarte? Ein übersichtsreferat über verschiedene praktische verschlusselungsmethoden. Nachrichten für dokumentation, 5:201-10 (1954).
52. Reference 51.
53. Washburn, Earle L. New York University. In: Baehne, G. W. Practical applications of the punched card method in colleges and universities; New York, Columbia University press, 1935, p. 135-44.
54. Reference 51.
55. National roster of scientific and specialized personnel, Report ... to the National resources planning board; Washington, D. C., U. S. Government printing office, 1942, p. 15-20.
56. Blashfield, W. H. Subscriber toll dialing tape reader. Electrical engineering, 72:786 (1953).
57. Richards, R. K. Arithmetic operations in digital

computers; Princeton, N. J., Van Nostrand, 1955, p. 178.
58. Reference 57, p. 178.
59. Reference 1, p. 48.
60. Reference 50, p. 283.
61. Brown, George B. Use of punched cards in acquisition work: experience at Illinois. College and research libraries, 10:219-20, 257 (1949).
62. Murphy, George M. An isotope file on punched cards. Journal of chemical education, 24:556-7 (1947).
63. Tagge, John. Hand-sort punch cards deliver maintenance data. Factory management and maintenance, 113 (no. 2):142-44 (Feb. 1955).
64. Young, George G. Borrower merely signs his name. Library journal, 74:12-16, p. 78 (1949).
65. Reference 1, p. 48-49.
66. Reference 50, p. 282.
67. Reference 1, p. 49.
68. Reference 1, p. 50-51.
69. Reference 1, p. 51-52.
70. Dyson, G. Malcolm. Some applications of the Dysonian notation of organic compounds. Journal of chemical education, 26:294-303 (1949).
71. Dunn, H. L. Adaptation of new geometric code to multiple punching in mechanical tabulation. American statistical association, Journal, 27:279-86 (1932).
72. Reagh, Russell R. A simplified code for multiple card punching. American statistical association, Journal, 29:182-83 (1934).
73. Reference 57, p. 178.
74. Reference 57, p. 179.
75. Reference 57, p. 179.
76. Reference 57, p. 179.
77. Reference 57, p. 179.
78. Reference 1, p. 63.
79. Reference 57, p. 179-180.
80. Engineering research associates, High-speed computing devices; New York, McGraw-Hill, 1950, p. 289.
81. Hartree, Douglas R. Calculating instruments and machines; Urbana, University of Illinois press, 1949, p. 59-60.

82. Royer, Elmer B. and Herbert A. Toops. The statistics of geometrically coded scores. American statistical association, Journal, 28:192-98 (1933).
83. Eckert, W. J. Electrons and computation. Scientific monthly, 67:315-23 (Jly.-Dec. 1948).
84. Reference 1, p. 53-54.
85. Reference 57, p. 184-192.
86. Reference 1, p. 54.
87. Reference 1, p. 54.
88. Reference 1, p. 55, 57.
89. Reference 50, p. 278.
90. Reference 57, p. 182.
91. Reference 50, p. 280.
92. Reference 50, p. 279.
93. Reference 50, p. 280.
94. Reference 57, p. 182.
95. Hamilton, F. E. and E. C. Kubie. The IBM magnetic drum calculator type 650. Association for computing machinery, Journal, 1 (no. 1):13-19 (Jan. 1954).
96. Reference 57, p. 184.
97. Photo-electric librarian. Electronics, 22:122, 158-66 (Sep. 1949).
98. Although Richards goes on to say: "However, the use of codes with more than 4 bits for the purpose of circuit simplification has not become widespread." See reference 57, p. 184.
99. Reference 57, p. 184.
100. Krider, L. D. Applications of automatic coding to small calculators. Eastern joint computer conference, Proceedings, 47:64-67 (Dec. 1954).
101. Reference 57, p. 185-192.
102. Reference 57, p. 185-192, passim.
103. Reference 57, p. 187.
104. Livesley, R. K. An introduction to automatic digital computers; Cambridge, Eng., University press, 1957, p. 16.
105. Reference 104, p. 16.
106. Reference 104, p. 16.
107. Reference 57, p. 5.
108. Reference 57, p. 5.
109. Reference 71.
110. Giffler, Bernard. Simplifying quality control op-

erations with marginal punch cards. Industrial quality control, 11 (no. 9): 20-23 (Je. 1955).
111. Reference 80, p. 84.
112. Reference 57, p. 184.
113. Reference 95.
114. Reference 57, p. 184.
115. Reference 1, p. 71.
116. Reference 1, p. 72.
117. Ruston, W. R. Die randlochkarte als hilfsmittel für die wissenschaftliche dokumentation. Nachrichten für dokumentation, 3:5-12 (1952).
118. Reference 51.
119. Reference 1, p. 71-78.
120. Jellinek, E. M., Vera Efron, and Mark Keller. Abstract archive of the alcohol literature. Quarterly journal of studies on alcohol, 8:580-608 (1947/48).
121. Reference 51.
122. Reference 1, p. 77-78.
123. Guy, A. G. and A. H. Geisler. A punch card filing system for metallurgical literature. Metal progress, 52:993-1000 (Jly.-Dec. 1947).
124. Moffit, Alexander. Punched card records in serials acquisition. College and research libraries, 7:10-13 (1946).
125. Wise, Carl S. A punched-card file based on word coding. In: Casey, Robert S. and James W. Perry, Punched cards; New York, Reinhold, 1951, p. 109.
126. Cox, Gerald J., C. F. Bailey, and Robert S. Casey. Punch cards for a chemical bibliography. Chemical and engineering news, 23:1623-26 (1945).
127. Reference 5.
128. Ames, Stanley R. and Wilma F. Kujawski. Use of punched cards for indexing and classifying biochemical literature. Special libraries, 39: 233-38 (1948).
129. Reference 126.
130. Reference 5.
131. Reference 1, p. 62.
132. Carr, A. M. Production control in a textile-finishing plant. In: Casey, Robert S. and James W. Perry, Punched cards; New York, Reinhold,

1951, p. 241.
133. Reference 1, p. 60.
134. Dyson, G. Malcolm. Punched-card systems and their application to library and technical work; I. Some applications of mechanical methods to library problems in organic chemistry. Association of special libraries and information bureaux, Report of the proceedings, 22:23-36 (1947).
135. Anderson, Isabella. The application of punch card filing in a chemical library. Illinois libraries, 31:406-11 (1949).
136. Westbrook, J. H. and L. H. DeWald. A modified punch card filing system for metallurgical literature. Metal progress, 54:324-27 (Jly.-Dec. 1948).
137. Reference 125, p. 111.
138. Reference 51.
139. Campbell, Kenneth N. and Barbara K. Campbell. Punched card code for general organic research. Industrial and engineering chemistry, 42:1458-60 (1950).
140. Reference 117.
141. Hood, T. A. Punch card for the field of metal finishing. Metal progress, 56:75-78 (Jly.-Dec. 1949).
142. Cox, Gerald J., Robert S. Casey, and C. F. Bailey. Recent developments in Keysort cards. Journal of chemical education, 24:65-70 (1947).
143. The coding capacity of each intersection in a triangular code may be determined by applying the permutation formula (4), where "n" is the number of rows and "m" is 2; the total number of intersections is determined by the combination formula (1), with "n" as the number of positions and "m" as 2. The total capacity is, therefore, the product of the results of these two computations; e.g., the 6-position, 2-row triangular code has a capacity of 2 x 15 = 30 items (cf. reference 1, p. 76).
144. Reference 23, p. 8-13.
145. Reference 23, p. 9, 11, 13.
146. Cochràn, S. W. Recent progress in patent classification. Industrial and engineering chemistry, 40:731-33 (1948).
147. Gull, Cloyd Dake. A punched card method for the

bibliography, abstracting, and indexing of chemical literature. Journal of chemical education, 23:500-07 (1946).
148. Nalbandjan, N. Le systeme à cartes perforées dans un service de bibliothéque. Les cahiers de la documentation, 2:43-47 (1948).
149. Reference 124.
150. Samain, Jacques. Une nouvelle méthode de sélection des documents: la mécano-mémoire. Hommes & techniques, no. 23/24:59-61 (Dec. 1946).
151. Telegraph. Encyclopaedia Britannica (1952 ed.), 21:887-93.
152. Reference 151.
153. Reference 57, p. 177.
154. Norris, Wells. How five-channel punched tape mechanizes office jobs. American business, 24 (no. 3):10-12, 36 (Mar. 1954).
155. Clark, Robert S. Books and reading for the blind; London, Library association, 1950, p. 7-14.
156. Reference 155, p. 10.
157. Reference 155, p. 8.
158. Formula (6) indicates that 64 combinations are possible; however, Braille did not consider the "no dot" arrangement as one of the combinations.
159. Typed figures translated into computer code. Engineering, 183 (no. 4749):348-49 (Mar. 15, 1957).
160. Shepard, David H. and Clyde C. Heasly, Jr. Photoelectric reader feeds business machines. Electronics, 28 (no. 5):134-38 (May 1955).
161. Printer plotter translates computer language. Electrical engineering, 76 (no. 3):262 (Mar. 1957).
162. Symbol generator and viewer. Computers and automation, 6 (no. 12):7-9 (Dec. 1957).
163. Soper, Alan K. Some observations on the use of punched cards for a personal information file. Aslib proceedings, 7:251-58 (1955).
164. Reference 125.
165. Mooers, Calvin N. Zatocoding applied to mechanical organization of knowledge. American documentation, 2:20-32 (1951).
166. Reference 163.
167. Reference 125.
168. Reference 163.

169. Mooers, Calvin N. Zatocoding for punched cards; Boston, Zator Co., 1950 (Zator technical bulletin, no. 30).
170. Reference 50.
171. Orosz, G. and L. Takács. Some probability problems concerning the marking of codes into the superimposition field. Journal of documentation, 12:231-34 (1956).
172. Reference 169, p. 18-19.
173. Reference 50, p. 293, 295.
174. Reference 50, p. 296.
175. Mooers, Calvin N. Putting probability to work in coding punched cards - Zatocoding; Boston, Zator Co., 1947? (Zator technical bulletin, no. 10).
176. Wise, Carl S. and James W. Perry. Multiple coding and the Rapid Selector. American documentation, 1:76-83 (1950).
177. Mooers, Calvin N. Coding, information retrieval, and the Rapid Selector. American documentation 1:225-29 (1950).
178. Wise, Carl S. Multiple word coding vs. random coding for the Rapid Selector. American documentation, 3:223-25 (1952).
179. Reference 177.
180. Reference 176.
181. Reference 177.
182. Reference 178.
183. Reference 176.
184. Reference 177.
185. New filing system developed for special collections. Library journal, 73:794-97 (1948).
186. Brown, William Fuller, Jr., and Glen Oneal, Jr. Library searches with punched-card machines. Science, 123:722-23 (Jan.-Je. 1956).
187. Rand Corporation. A million random digits with 100,000 normal deviates; Glencoe, Ill., Free press, 1955.
188. Reference 186.
189. Reference 1, p. 196-98.
190. Reference 125.
191. Reference 169, p. 9.
192. Reference 50, p. 298-99.
193. Code chemicals for index. Science news letter,

54:39 (Jly.-Dec. 1948).
194. Sherman, Jack. The use of four-hole randomly punched cards for abstracting publications and reports into IBM cards. Special libraries association, Texas chapter, Bulletin, 8 (no. 4):9-14 (May 1957).
195. Schultz, Claire K. Coding literature on punched cards; unpublished MLS thesis, Drexel Institute of Technology, 1952.
196. Sebring, M. W. A marginal punched card system for a specialized information collection. American documentation, 4:18-22 (1953).
197. Perry, James W. Superimposed punching of numerical codes on hand-sorted punch cards. American documentation, 2:205-12 (1951).

Index

N

V

W

X

Y

Z